Evolutionary Paths towards Multicellularity and Cellular Differentiation: A Study of Developmental Strategies in Cyanobacteria

Dissertation

zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von
Valentina Rossetti
aus
Italien

Promotionskomitee:

Prof. Dr. Homayoun Bagheri (Vorsitz)
Prof. Dr. Martin Ackermann
Prof. Dr. Andrew Barbour
Prof. Dr. Sebastian Bonhoeffer
Prof. Dr. Andreas Wagner

Zürich, 2012

Evolutionary Paths towards Multicellularity and Cellular Differentiation: A Study of Developmental Strategies in Cyanobacteria

Dissertation

zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

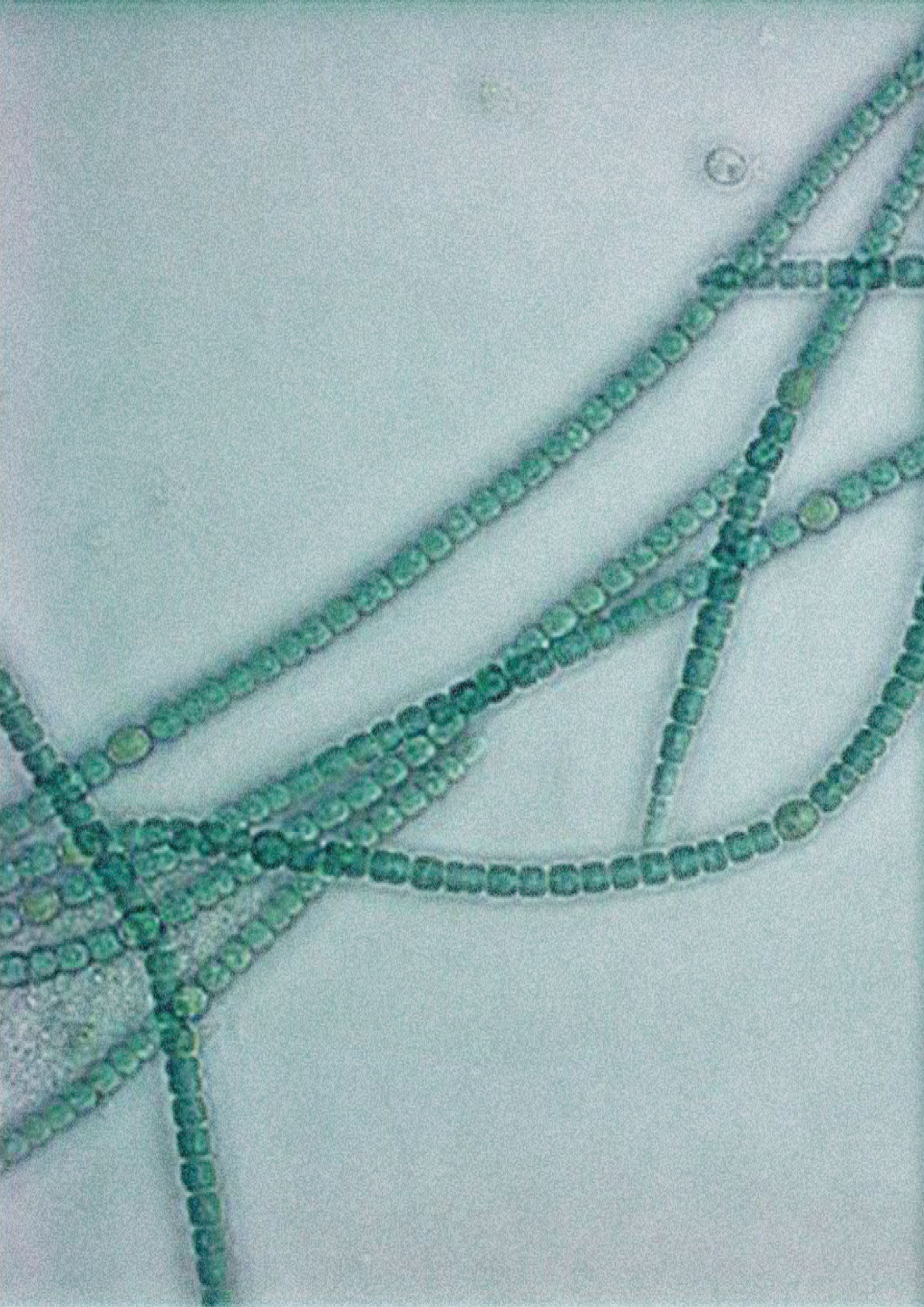
von
Valentina Rossetti
aus
Italien

Promotionskomitee:
Prof. Dr. Homayoun Bagheri (Vorsitz)
Prof. Dr. Martin Ackermann
Prof. Dr. Andrew Barbour
Prof. Dr. Sebastian Bonhoeffer
Prof. Dr. Andreas Wagner

Zürich, 2012

Die vorliegende Arbeit wurde von der Mathematisch-naturwissenschaftlichen Fakultät der Universität Zürich im FS semester 2012 als Dissertation angenommen.

Promotionskomitee:
Prof. Dr. Homayoun Bagheri (Vorsitz)
Prof. Dr. Martin Ackermann
Prof. Dr. Andrew Barbour
Prof. Dr. Sebastian Bonhoeffer
Prof. Dr. Andreas Wagner



Abstract

This thesis, developed at the interface between biology and applied mathematics, explores fundamental steps in the evolution of life, namely the evolution of multicellularity and cellular differentiation. Cyanobacteria, the main model organism for this work, were present on Earth billions of years ago and are a living image of the primordial emergence of multicellular organisms. The three themes of this work aim to study the evolutionary and ecological conditions governing this emergence.

In chapter 2, we developed a theoretical model and tested it against experiments, showing that multicellular life cycles can emerge as a result of birth/death processes in an ecological context. When cell birth and death rates of multicellular filamentous bacteria are dependent on the density of cells in a population, a predictable cycle between short and long filament lengths is produced. Changes in generation time in bacterial populations can alter length distributions, and even when fitnesses do not differ, average filament lengths can differ drastically. This shows that differences in fitness are not the sole explanation for the evolution of multicellular life cycles.

Chapter 3 is devoted to investigate how differentiation would evolve in a unicellular context. This would require collaborative consortia of single cells that survive through the exchange of common goods. Our mathematical model of cyanobacterial species shows that in such a configuration, deleterious mutations producing cheater cells would lead to the collapse of the system. Moreover, an optimal ratio between cell types that ensures an optimal growth can be achieved only in multicellular organisms. These findings indicate multicellularity as a necessary condition for the stability and the optimization of division of labor in cyanobacteria.

Through terminal differentiation, cyanobacteria achieve a spatial separation of photosynthesis and nitrogen fixation, that are chemically incompatible. Undifferentiated species separate the tasks in time according to a circadian rhythm. In chapter 4, we use a theoretical model to compare the biomass production of the two species types at different latitudes. We find that an optimal resource investment into reproduction and nitrogen fixation can enhance the biomass production of terminally differentiated species as if there were no constraints.

The results in this thesis offer new perspectives on the evolution of multicellularity and cellular differentiation in bacteria and higher organisms.

Zusammenfassung

Diese Dissertation, an der Schnittstelle zwischen Biologie und angewandter Mathematik entwickelt, erforscht grundlegende evolutionäre Schritte, nämlich die Evolution der Mehrzelligkeit und der zellulären Differenzierung. Cyanobakterien, die hier als Modellorganismus wirken, fanden sich auf der Erde schon vor Billionen Jahren ein, deshalb sind sie ein lebendiges Bild der ursprünglichen Entwicklung der mehrzelligen Organismen. Die drei Themen dieser Arbeit haben das Ziel, auf die aufbauenden, evolutionären und ökologischen Bedingungen für solche Entwicklungen zu erforschen.

In Kapitel 2 haben wir ein theoretisches Modell entwickelt und es mit Experimenten geprüft. Damit zeigen wir, dass mehrzellige Lebenszyklen als Konsequenz des Geburts- und Todesprozesses der Zellen in einem ökologischen Zusammenhang entstehen können. Wenn Zellgeburts- und Todesrate der mehrzelligen fadenförmigen Bakterien von der Populationsdichte abhängig sind, wird ein voraussagbarer Zyklus zwischen kurzen und langen Filamenten erbracht. Generationszeitsveränderungen in Bakterienpopulationen können die Längeverteilung modifizieren. Auch wenn die Fitness der Populationen nicht abweicht, kann die durchschnittliche Fadenlänge sehr unterschiedlich sein. Das zeigt, dass Unterschiede in Fitness nicht die einzige Erklärung für die Entstehung von mehrzelligen Lebenszyklen sind.

Kapitel 3 forscht nach, wie Differenzierung in einem einzelligen Zusammenhang sich entwicklen könnte. Das würde zusammen arbeitende Konsorte von einzelligen Organismen voraussetzen, die durch den Umtausch von gemeinsamen Gütern fortleben. Unser mathematisches Modell von Cyanobakterien zeigt, dass in einer solchen Umgebung schädliche Mutationen, die Betrüger-Zellen produzieren, den Zusammenbruch des Systems verursachen würden. Zudem ist ein optimales Verhältnis zwischen Zelltypen, das ein optimales Wachstum bewirkt, nur in einer mehrzelligen Umgebung möglich. Diese Ergebnisse zeigen, dass Mehrzelligkeit eine notwendige Bedingung ist für die Stabilität und die Optimierung der Arbeitsteilung zwischen den Zellen.

Durch terminale Differenzierung erlangen Cyanobakterien eine räumliche Trennung von Photosynthese und Stickstofffixierung, die chemisch inkompatibel sind. Undifferenzierte Arten trennen die Aufgabe zeitlich, einem Tagesrhythmus gemäss. In Kapitel 4 benutzen wir ein theoretisches Modell um die Biomasseproduktion beider Arten an verschiedenen geografischen Breiten zu vergleichen. Wir finden

heraus, dass ein optimaler Mitteleinsatz im Bereich der Vermehrung und der Stickstofffixierung die Biomasseproduktion der differrenzierten Arten erhöhen kann, als ob es keine Einschränkung gäbe.

Die Ergebnisse dieser Dissertation zeigen neue Perspektiven auf in der Evolution der Mehrzelligkeit und der Differenzierung in der Bakterienwelt und in höheren Organismen.

Laudato sii, mio Signore, per tutte le tue creature
(San Francesco d'Assisi)

Contents

1 Preface and thesis overview

1.1 Preface

During the course of my PhD, I had the opportunity to start understanding how fundamental and concrete is the role of bacteria in the evolution of life on our planet. Bacteria made Earth a habitable planet and keep contributing to its life. The study of their evolutionary history is of fundamental help in understanding the development of more complex organisms. The following pages illustrate how the developmental strategies of cyanobacteria represent milestones in the evolution of multicellularity and cellular differentiation.

1.1.1 Multicellularity and cellular differentiation

Considering that the human body contains on average several tens of trillions of cells, there is no doubt that we are multicellular individuals. Among these cells, one can distinguish many different cell types, hence we are differentiated multicellular organisms. The question of how and why multicellularity could evolve and give rise to increasingly complex life forms has been tackled in several fields. Similarly, the mechanistic aspects of the development of differentiated cells has been an intriguing topic for many scientists. The evolutionary roots of such fundamental biological processes date back to billions of years ago, when the first organisms developed on the primordial planet. According to fossil records, bacteria already exhibited multicellular shapes more than 3200 million years ago [173].

A wide range of factors have been hypothesized to lead to the evolution of multicellularity and division of labor. Reduced risk of predation, enhanced motility, development of specific strategies for dispersion and allowance for coexistence of chemically incompatible processes are just some of the most important ones

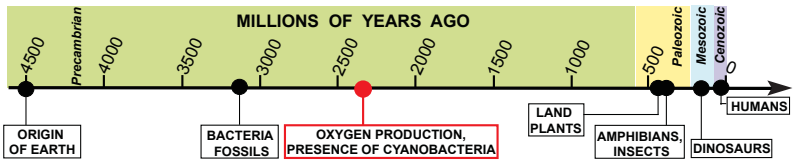


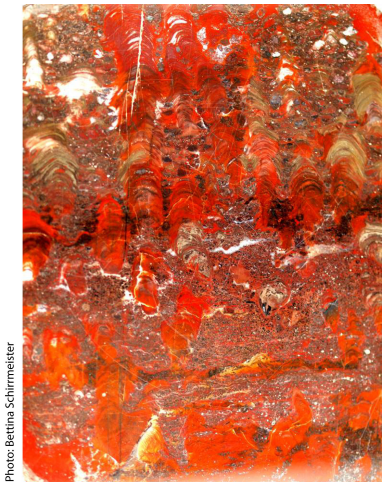
Figure 1.1: On the time line the most relevant evolutionary stages are indicated. Cyanobacteria were most probably present already 2.5 billion years ago, when the great oxygenation took place.

[17, 86]. Some extant species of myxobacteria and cyanobacteria, as well as eukaryotes such as the volvocales, retain simple multicellular traits and are hence excellent model organisms for the study of the evolution of multicellularity [93, 217], which arose independently several times in different kingdoms and phyla [17, 18, 95].

The subsequent chapters in this thesis include detailed reviews of the specific topics related to multicellularity and differentiation. As cyanobacteria are the model species in this thesis, the next section is devoted to present the role of cyanobacteria in microbial evolution.

1.1.2 Cyanobacteria in microbial evolution

During Darwin's time, the bacterial world was still mostly undiscovered, although a pioneer of microbiology, Antoni van Leeuwenhoek (1632-1723), had already attested to the existence of microorganisms. Darwin did not consider bacteria in his Origin of Species. However, if he could have seen even some of the precambrian fossil records discovered in modern times, he would have included them in his tree of life. Bacteria represent one of the three domains of life (Bacteria, Archea and Eucarya) and can illustrate some of the very first steps in the history of life. Cyanobacteria occupy a special place in this history. Some scientists refer to the the proterozoic era (2500-570 Ma) as the Age of cyanobacteria [174]. From the mid 1960s onwards, advances in precambrian paleobiology contributed to the discover of many fossils belonging to the proterozoic era. Among them, cyanobacteria are the most abundant species, and can be traced back to at least 2-2.5 billions years ago [171, 22, 175] (Figure 1.1). Often, filamentous shapes and chains of cells imprinted on rocks have a high resemblance to the extant species of multicellular



a

b

Figure 1.2: A 2 Ma old stromatolite. The red color is due to iron oxidation. (b) Stromatolites in Yalgorup National Park (Australia). In this environment, the growth rate of the stromatolites is about 0.4 mm per year.

cyanobacteria [172]. An additional type of fossil is represented by the stromatolites (Figure 1.2a,1.2b). These are lithic layer formations, typically found in shallow water, in which sediments and bacteria are trapped and stratified together. At present, one of the best known examples of stromatolites is in Australia (Figure 1.2b). The morphology of cyanobacteria has been almost unaltered for billions of years, indicating that these organisms possessed characteristic that allowed them to survive after many environmental changes on our planet.

Cyanobacteria are photoautotrophic microorganism living in almost any aquatic and terrestrial environment. Originally called blue-green algae, their name derives from a blue pigment, phycocyanin, used to absorb light. Cyanobacteria possess a complex photosynthetic apparatus, whereby they use solar light for oxygenic photosynthesis. Their photosynthetic mechanism is based on two photosystems (photosystems I and II, [82, 113]). According to the endosymbiotic theory, plant chloroplasts developed from cyanobacteria that had the photosynthetic machinery [62, 150]. Although the origin of oxygenic photosynthesis is under debate [23, 22, 97], there is evidence that cyanobacteria contributed to the great oxygenation event around 2.4 billion years ago.

Many cyanobacterial species are able to fix molecules of N_2 present in the envi-

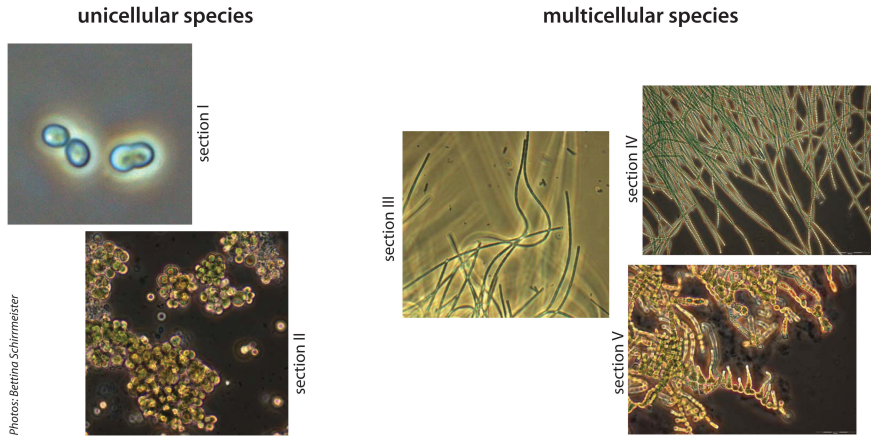


Figure 1.3: Examples of cyanobacterial species according to the classification of Rippka [152].

ronment through nitrogen fixation (diazotrophs). This process takes place in cells that express the nitrogenase enzyme. However, nitrogenase is oxygen-sensitive, whereby oxygen, a byproduct of photosynthesis, inhibits nitrogen fixation. As it will be explained later, the need for a separation of these incompatible processes led cyanobacteria to adopt different developmental strategies, ranging from a circadian rhythm to cellular differentiation [57, 12].

With more than 2500 recognized species (source: AlgaeBase), cyanobacteria exhibit a great diversity (Figure 1.3). The most used classification is due to Rippka [152] and divides them in five sections according to their morphology, cellular differentiation and cell division mechanism. They range from unicellular (sections I and II) to filamentous (sections III and IV) and branching filamentous species (section V). Species belonging to the first three clades are undifferentiated, while those in sections IV and V exhibit cellular differentiation.

1.1.3 Relevance of Cyanobacteria

Cyanobacteria are not only relevant to evolutionary biology, and are also well known for other aspects. Some species are toxic for animals and humans, as they

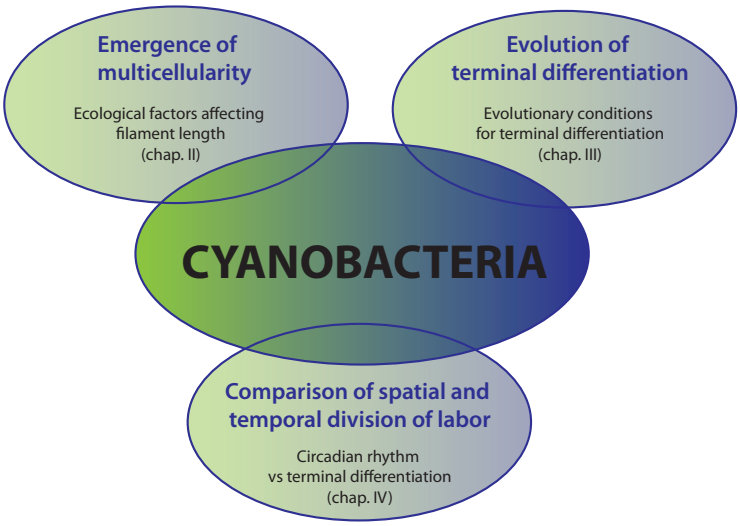


Figure 1.4: Schematic view of the thesis. Structured in three themes, it aims to understand the evolutionary, ecological and physiological conditions leading to the development of multicellular individuals and differentiated cells in cyanobacteria.

can produce hepatotoxins, neurotoxins and toxic alkaloids [28]. People and animals can be severely affected by cyanotoxins by drinking or bathing in contaminated water [29, 80]. Cyanobacterial blooms in seas and oceans can be especially dangerous, although they are observed mainly for a restricted number of species and more frequently in particular environments and seasons [176].

On the other hand, cyanobacteria can be used for bioremediation and wastewater treatment. Being photoautotrophic and diazotrophic, they represent a valid alternative to heterotrophic species [101, 34]. Recently, applications in agro-industrial wastes and wastewaters have been developed [116].

An additional application of cyanobacteria is their use as food supplements. Since some active components of cyanobacteria are perceived to have beneficial effects on human health, some species have been used in dietary products [89]. However, their long term effects are not known [58].

1.2 Thesis overview

The study of the factors leading to the evolution of cyanobacterial developmental strategies implies the consideration of different points of view and a multidisci-

plinary approach. With this perspective, the present work combines mathematical models with experimental tests. Figure 1.4 illustrates how the evolution of multicellularity and cellular differentiation in cyanobacteria is approached from different angles in this thesis. We first frame density dependent growth of multicellular filaments in the context of classical ecological theory. The mathematical model is complemented by morphological measurements on experimental filamentous species (Chpt. 2). We then consider the specific case of cyanobacteria (Chpt. 3). With the help of a mathematical model of extant and alternative hypothetical forms of cellular differentiation, we investigate the conditions allowing for the evolution of cyanobacteria towards multicellularity and division of labor. The study of cellular differentiation in cyanobacteria is extended by modeling two types of division of labor, namely spatial and temporal (Chpt. 4). A comparison of physiological performance in response to variation in environmental and physiological parameters is used to obtain insights into the evolutionary advantages of each species type.

As a further development, we present plans to study differentiation mechanisms in *Trichodesmium* (Appendix D).

1.2.1 Emergence of multicellular life cycles (Chpt. 2)

As already mentioned, filamentous bacteria are among the oldest and simplest multicellular life forms known. In Chpt. 2 we use computer simulations combined with experiments that address cell division in filamentous bacteria. In this context we investigate some of the ecological factors that can lead to the emergence of multicellular life cycles. Using the classical ecological assumption that birth and death rates are density dependent, the effect of this assumption on filament length dynamics is investigated. The model has been tested on heterotrophic and cyanobacterial species. In order to make this new framework comparable to experimental results, a new formulation of generation time in natural bacterial populations is developed.

The model predicts that if cell division and death rates are dependent on the density of cells in a population, a predictable cycle between short and long filament lengths is produced. During exponential growth, there will be a predominance of

multicellular filaments, while at carrying capacity the population converges to a predominance of short filaments and single cells. Model predictions are experimentally tested and confirmed in cultures of heterotrophic and photoautotrophic bacterial species. Thanks to our new formulation for generation time, it is shown that changes in generation time can alter length distributions. The theory predicts that given the same population growth curve and fitness, species with longer generation times have longer filaments during comparable population growth phases.

Characterization of the environmental dependence of morphological properties such as length, and the number of cells per filament, helps in understanding the pre-existing conditions for the evolution of developmental cycles in simple multicellular organisms. Moreover, the theoretical prediction that strains with the same fitness can exhibit different lengths at comparable growth phases has important implications. It demonstrates that differences in fitness attributed to morphology are not the sole explanation for the evolution of life cycles dominated by multicellularity.

1.2.2 The evolution of multicellularity and terminal differentiation (Chpt. 3)

A common trait often associated with multicellularity is cellular differentiation, which is a separation of tasks through the division of labor. In principle, the division of labor does not necessarily have to be constrained to a multicellular setting. In Chpt. 3 we focus on the possible evolutionary paths leading to terminal differentiation in cyanobacteria. We develop mathematical models for two developmental strategies. One, of populations of terminally differentiated single cells surviving by the exchange of common goods. Second, of populations exhibiting terminal differentiation in a multicellular setting. While the latter corresponds to extant heterocystous species, the former has no correspondence to natural populations. We hence investigate the conditions that make this kind of differentiation scarce. We first test the ability of the two strategies in surviving against disruptive mutations (i.e. “cheaters”) and in optimizing the ratio of vegetative to heterocystous cells. In addition, we compare the performance of differentiated populations to undifferentiated ones that temporally separate tasks in accordance to a day/night

cycle. We then validate some predictions of our model by deriving phylogenetic relationships obtained from analyzing 16S rDNA sequences of different cyanobacterial strains. In line with studies indicating that group or spatial structure are ways to evolve cooperation and protect against the spread of cheaters, we find that compartmentalization afforded by multicellularity is required to maintain the vegetative/heterocyst division in cyanobacteria. Moreover, multicellularity allows for selection to optimize the carrying capacity. The phylogenetic analysis are in accordance with the theoretical results, indicating that terminally differentiated cyanobacteria evolved after undifferentiated species. In addition we show that in regimes of very short daylight periods, terminally differentiated species perform worse than undifferentiated species that follow the day/night cycle, indicating that undifferentiated species have an evolutionary advantage in regimes of short daylight periods.

1.2.3 Comparison of spatial and temporal differentiation (Chpt. 4)

As mentioned in the introduction, diazotrophic cyanobacteria are able to perform nitrogen fixation. However, the inhibition of nitrogen fixation due to oxygen produced by photosynthesis imposes biochemical constraints on these species. In order to overcome these constraints, cyanobacteria evolved two main developmental strategies. Undifferentiated species adopt a circadian rhythm to achieve temporal separation. Terminally differentiated heterocystous species opt for spatial separation. In the differentiated species, the specialized heterocystous cells serve as a microoxic environment in which nitrogen fixation can take place without being inhibited.

The mathematical model presented in Chpt. 4 is designed to assess the advantages of each species type in terms of biomass production and population size. Parameters affecting the model are geographical position and cell investment rates in cellular division and nitrogen fixation. Moreover, for purposes of comparison, we also model an idealized species in which nitrogen and carbon fixation occur without biochemical constraints. On the basis of real solar irradiance data at different latitudes, we simulate the population dynamics over a period of 40 years

for the three mentioned species types: circadian, heterocystous and the idealized species.

Our results show that regardless of latitude, the biomass accumulation of optimized heterocystous cyanobacteria is comparable to that of the idealized unconstrained species. This indicates that division of labor overcomes biochemical constraints and enhances biomass production. The circadian rhythm modeled here is subject to a strict temporal task separation that hinders a high biomass production. This would suggest that selection is expected to reduce nitrogenase sensitivity to oxygen in such species. In addition, we show the existence of a trade-off between population size and biomass accumulation, whereby each species can optimally invest resources to be proficient in biomass production or population growth, but not necessarily both. Finally, the model produces chaotic dynamics for population size, which can be relevant in the study of cyanobacterial blooms.

2 Emergent multicellular life cycles in filamentous bacteria owing to density-dependent population dynamics

Valentina Rossetti, Manuela Filippini, Miroslav Svercel, A.D. Barbour and Homayoun C. Bagheri

[*Journal of the Royal Society Interface*, (2011) 8, 1772−1784]

2.1 Introduction

Multicellularity is an organizational characteristic present in the majority of organisms whose size surpasses microbial scales. Phylogenetic inference suggests that evolutionary transitions to multicellularity have occurred several times during the history of life [26, 17, 32, 92, 64, 156]. Nonetheless, despite its fundamental importance, it is difficult to empirically study the evolutionary and ecological forces that may lead to a transition from single-celled to multicellular organization. One set of theoretical explanations are based on examining the consequences of a shift between units of selection. These approaches consider the changing boundaries of an individual after a transition to multicellularity, and how the different activities of component cells can potentially lead to synergies that increase the fitness of the multicellular assemblage [26, 183, 120, 64]. Experimental results concerning social evolution in microbes are also largely compatible with this perspective [24, 194, 207, 148, 63, 1].

An additional set of explanations for multicellularity are based on the potential selective advantages associated with increased size. Among the proposed size-related advantages are increases in the efficiencies of feeding [39, 16, 160, 141, 98, 186, 11], improved dispersal [16, 187, 218], and predator avoidance [193, 65, 179, 66, 140]. Though many of the proposed explanations are highly plausible, there is a lack of empirical validation. Some notable exceptions are measurements of motility and feeding efficiencies in Volvocalean green algae [186, 187], and density dependent growth in *Myxoccoccus xanthus* [160, 178, 11].

Multicellular bacteria with a filamentous form are ubiquitous in many environments, including freshwater, oceans, soil, extreme habitats and the human body [216, 136, 199, 159]. Extensive empirical work has been done to examine and monitor filamentous bacteria that can be toxic or problematic in the environment [85, 176, 149, 128]. Some filament-forming cyanobacteria also develop specialized terminally differentiated cells, named heterocysts, that fix nitrogen and allow for the division of labor [192, 220]. Though differentiation can represent a clear evolutionary advantage, theoretical and phylogenetic evidence suggests that in the cyanobacterial case, undifferentiated multicellularity evolved prior to differentiated multicellularity [161]. In aquatic bacteria, the most common hypothesis

for the advantage of filamentation —and hence undifferentiated multicellularity— is an increase in size and the concomitant defense against predation by grazers [65, 179, 83, 139, 66, 138]. However, experiments indicate that the avoidance of predatory grazers is not the only factor causing an increased frequency of filamentous bacteria in aquatic environments [222]. Notably, in some bacterial species, filament formation appears to be dependent on the growth state of the population, whereby an increase in the dilution rate of chemostat cultures leads to longer filament lengths [66, 67].

No theoretical studies address the distribution of filament lengths and the population dynamics leading to shorter or longer filaments, although differences in length can reflect the extent to which a species is able to maintain multicellularity. Environmental conditions such as temperature, solar irradiation and nutrient concentrations have been detected as factors in determining the mean size (filament length) of different cyanobacterial species [87, 221, 100]. Several of these factors also contribute to competition between species and adaptation to different niches [106, 131]. Filament breakage can occur because of external mechanical stress, lytic processes initiated by pathogens [180, 206, 211] or programmed cell death [36, 110, 130, 14, 3].

In the present study, we deliberately avoid a modelling framework in which one assumes an *a priori* fitness advantage to multicellularity. We also do not aim to provide mechanistic explanations for the "origin" of multicellularity. Rather, by considering cellular growth and division in a filamentous context, we investigate the constraints that would affect filament formation and the maintenance of multicellularity. This gives an indication of the capability of bacterial organisms to evolve multicellular life cycles according to their life history traits. Our model is built on the basic assumption that the growth of a population of filamentous bacteria and the changes in length of the filaments can be set in an ecological framework. In classical population dynamics, the change in population size is governed by the processes of birth and death (and sometimes migration), combined in the well known logistic equation due to Verhulst. In this model, birth and death rates are usually assumed to be decreasing and increasing functions of the population density, respectively. When the birth and death rates are the same, the population density is at a steady state and is said to be at its carrying ca-

pacity. The value at which the two rates are equal will be here referred to as the turnover rate. The same carrying capacity can be achieved with different birth and death rate functions, and therefore at different corresponding turnover rates (Figure 2.1a).

Our approach to study distributions of bacterial filament length places the described density dependent concepts in the context of a population of cells that can form filamentous individuals. Although prokaryotes had been commonly thought to be in principle immortal [219, 158], there is now evidence for programmed cell death and lysis [151], and in addition aging and senescence [2]. Taking this into account, a multicellular filament is made up of several cells, and a population of filaments can be also considered as a population of cells. In this case, the population size is given by the total number of cells, while the birth and death rates coincide with the birth and death rates of each cell of the filaments. In such a population, one can vary the turnover at which a given fixed carrying capacity is achieved and ask the question if different turnovers correspond to different growth strategies; namely whether during its growth cycle the population is composed mostly of a few long filaments, or of many short filaments. Such a line of questioning is possible because as it will be shown, the rate at which a population reaches its carrying capacity can be different from the rate at which it reaches its stationary filament length distribution. The present work is an attempt to model and experimentally test the dynamics of filamentous bacteria by only considering basic life history traits of the population. The distribution of filament lengths in the different growth phases can provide insights into the developmental strategies of different filament-forming species, and the population conditions influencing multicellular development.

In the following sections, we first state the model assumptions and we specify the experimental setup for the culture of five filamentous bacterial species. Since turnover is a key factor in our model, we chose species that also exhibit different turnovers. We hence cultured heterotrophic species with fast generation time and high turnover, and photoautotrophic species characterized by a slower generation time and a low turnover. Heterotrophs differ from photoautotrophs in the way they obtain organic carbon for growth. While the former absorb organic compounds from the environment, the latter autonomously fix carbon using solar light as an

energy source. We present the simulation results and test the model through a comparison with our experimental results.

2.2 Methods

2.2.1 Theoretical model

We consider a population of multicellular, undifferentiated, non-branching filamentous bacteria, whose growth is regulated only by the density-dependent birth and death rates. In this context, the total number of cells of the population is used as the population size. The birth and death rates are represented by birth and death probabilities per iteration per cell. At every iteration, each cell in a filament has a certain probability of dividing into two daughter cells or of dying. Based on these probabilities, the number of cell divisions and cell lyses are calculated in each filament. Accordingly, the filaments are elongated and then broken into smaller pieces. We simulate the change in time of the population until it has stabilized at its carrying capacity. At each iteration, the following steps are performed (see Figure A.1):

(i) Computation of the total number of cells in the population $N_c = \sum_{i=1}^{L} i\, n_i$, where L is the maximal length of the filaments in the current generation and n_i is the number of filaments of length i.

(ii) Computation of the birth and death probabilities per cell per iteration, corresponding to the birth and death rate functions, respectively $\beta = \beta(N_c)$ and $\delta = \delta(N_c)$:

$$\beta(N_c) = -c_1 N_c + 1, \qquad c_1 = \frac{1-\theta}{N_c^*} \tag{2.1}$$

$$\delta(N_c) = c_2 N_c, \qquad c_2 = \frac{\theta}{N_c^*} \tag{2.2}$$

where N_c is the total number of cells and N_c^* the carrying capacity. The latter is defined as the value of the population size such that $\beta(N_c^*) = \delta(N_c^*) = \theta < 1$, where θ is the turnover. The birth rate is a decreasing function of N_c. The death rate is an increasing function of N_c.

(iii) For each filament, we compute the number of births b and number of deaths d occurring in the filament. As β and δ are the per cell probabilities, in a filament of length i, the total number of events b or d can be viewed as the number of successes in i trials, each one with probability of success β or δ. By picking from a binomial distribution B, we determine b and d for each filament of length i:

$$\begin{aligned} b &= \text{random pick from } B(i,\beta) \\ d &= \text{random pick from } B(i,\delta)\,. \end{aligned}$$

The filament is elongated by b cells. Then, a number of d cells computed on the basis of the original filament length i are randomly selected for lysis in the elongated filament. The filaments resulting from the breakages are stored for the next generation. After completing the scan of all the filaments, the new population is set for the next iteration. The choice of initial conditions and the order at which birth and death are applied to filaments do not significantly affect the results of the simulations. The default initial condition for the simulations is a single cell. Alternative algorithms implementing a reverse or a random order of birth and death steps have been tested and produced similar results (figures not shown).

As stated in point (ii), this model implements linear birth and death rate functions, with fixed carrying capacity. Alternatively, nonlinear functions have been implemented and provided qualitatively comparable results (see § A.2). Considering the difference between those rates, given by $\beta - \delta = -(c_1 + c_2)N_c + 1$, one can observe that the net growth rate $(c_1 + c_2)$ is independent of the turnover θ, since $c_1 + c_2 = 1/N_c^*$. This implies that regardless of the turnover, the population size is expected to grow with the same curve for any turnover.

2.2.2 Relation between turnover and generation time

Before presenting the results, we provide a formula that helps to understand the connection between the model parameters and the experimental data. In the theoretical model, the turnover is the characterizing property of a strain. However, in natural bacterial populations, turnover is not an easily measurable quantity. For bacteria, doubling time during the exponential growth phase is usually calculated in lieu of generation time. Two populations with the same growth curve also have

the same doubling time. Nonetheless, the reality is that the internal dynamics leading to this growth curve can vary. In the same time period, the same concentration of bacteria can be achieved by undertaking many births and deaths, or by conversely having less births and also less deaths. This leads to different turnovers and generation times. Moreover, at carrying capacity the population density is stable, hence there is no indication of generation time at that stage by measuring doubling time. Here we calculate the generation time of bacterial populations at carrying capacity in relation to turnover rate.

According to a common approach in ecology to describe the dynamics of an age-structured population [109], the mean age of parents at childbirth in a population with stable age distribution is calculated as:

$$G = \big(\sum_{x=0}^{\infty} x l_x m_x\big) \big/ \big(\sum_{x=0}^{\infty} l_x m_x\big) \tag{2.3}$$

where x is the age, l_x is the probability of survival from conception to age x, and m_x is the mean number of offspring at age x. G can be also interpreted as a measure of generation time without making any assumption about the growth rate of the population [33]. In our case, age is measured in number of iterations. We consider bacterial cells as individuals that give birth by cell division and die by a lytic process. Suppose that a cell survives a given iteration with probability $1-\delta$. Then the probability that a cell is living after x iterations if δ remains constant is given by $l_x = (1-\delta)^x$. If also the probability of cell division is β at every iteration, the average number of offspring at a given age x coincides with the probability of birth β, whereby $m_x = \beta$. The generation time of a cell can then be expressed as

$$G_{cell}(\beta, \delta) = \big(\sum_{x=0}^{\infty} x(1-\delta)^x \beta\big) \big/ \big(\sum_{x=0}^{\infty} (1-\delta)^x \beta\big) = \frac{1-\delta}{\delta}\,, \tag{2.4}$$

where the last equivalence was obtained by using geometric series. The standard formula in (2.3) applies when the growth rate is constant. Given the logistic nature of our model however, the functions β and δ are actually dependent on the population size N_c. Only when the population approaches the carrying capacity, namely when $N_c \to N_c^*$, we have that $\beta, \delta \to \theta$, hence the growth rate is constant

(equal to 0). We then present a formula for the generation time at the carrying capacity. Substituting $\beta = \delta = \theta$ in eq. (2.4), the generation time is given by

$$G_{cell}(\theta, \theta) = \frac{1-\theta}{\theta} \tag{2.5}$$

which is a decreasing function of the parameter θ (Figure A.2). Hence, a population with low turnover has a long generation time, whereby cell birth and deaths occur at a low rate. On the other hand, populations with high turnover have a short generation time, whereby births and deaths occur rapidly.

2.2.3 Experimental setting

Bacterial Strains. The heterotrophic strains used in the experiments were *Rudanella lutea* DSM 19387^T [213] and two new bacteria isolated from a mud sample from tidal flats in Fedderwardersiel, on the North Sea coast of Germany, *Fibrella aestuarina* BUZ 2^T [49] and strain *Fibrisoma limi* BUZ 3^T (Filippini et al., in press). The autotrophic strains were two axenic cyanobacteria, *Nostoc muscorum* SAG 25.82 (identical strains ATCC 27893; PCC 7120) and *Anabaena variabilis* SAG 1403-4b (identical strains CCAP 1403/4B; UTEX 377; ATCC 29211; PCC 6309).

Batch culture experiments for heterotrophs. The growth of the three bacterial species was monitored in batch cultures. Stock bacteria were grown on R2A plates for 3-4 days before starting the experiment. Colonies were homogenized in 0.7% NaCl and finally distributed into 50-ml flasks containing 25 ml of SM (DSMZ 7) or R2A medium (DSMZ 830) for *F. limi*, *R. lutea* and *F. aestuarina*, respectively reaching a starting optical density (OD) of 0.04. Bacteria were grown on a shaker (120 rpm) at 29℃ and after different time periods sub-samples were analyzed for OD and filament length. For *R. lutea* and *F. aestuarina*, two batch cultures (A & B) were prepared and analyzed, whereas for *F. limi* only one batch culture (A) was used for the experiment. Additionally, for each batch, three technical replicates (i.e. a1, a2, and a3) were made. The OD was measured in a spectrophotometer (SpectraMax 384 Plus, Molecular Devices, USA) at 700 and 600nm. The samples for length measurement were fixed with formaldehyde (final

concentration, 2%), 20μl were distributed on a microscope slides and pictures were taken with a digital camera (Color View, Soft Imaging System, 10x, 20x and 40x objective) connected to a phase-contrast microscope (Olympus BX 51, Germany). For each time point, between 10 and 20 pictures from separate fields were taken and the sizes of the filaments were individually determined by hand using the Soft Imaging software CellF (Olympus, Germany). At least 300 filaments from 10 different pictures were counted for each time point, in order to have a representative distribution of filament lengths.

Batch culture experiments for cyanobacteria. As in the heterotrophic case, the growth of the two cyanobacterial species was monitored in batch cultures. Stock cultures were grown in 20 ml of BG11 medium [152] in 50 ml Erlenmeyers one week prior to the experiment. Two ml of grown culture were added to 18 ml of BG11 medium and cultivated on a shaker (150 rpm) at room temperature and constant light (670 lux) for three weeks. Every two days, 200 μl subsamples were taken, transferred into a microtiter plate and analyzed for OD (at 430, 630 and 680 nm; SpectraMax 384 Plus). Filament length was measured by distributing 20 μl samples in a counting chamber Neubauer-improved (Paul Marienfeld GmbH & Co, Germany) and pictures were taken with a digital camera connected to a phase-contrast microscope. Approximately 18 pictures with a 10x magnification objective (or 2 pictures with a 4x magnification objective) were taken for each time point, and the sizes of the filament were determined using the Soft Imaging software CellF. For each time point, at least 50 filaments were counted in order to have a representative distribution of filament lengths. Three experiment replicates (Erlenmeyer flasks) of each cyanobacteria and four measurement samples of each replicate were analyzed for OD. Initial stock cultures were also examined. For filament lengths, four measurement samples were assayed from a single replicate. In Figures 2.3d-e, we show the results of two experiment replicates (experiments A and B).

To extrapolate the cell number from the filament length, an estimated size of cells forming a filament was measured from several light microscopy pictures. From at least 60 observations, the estimated mean cell lengths are 6.0 μm, 6.7 μm and 5.9 μm for *R. lutea*, *F. aestuarina* and *F. limi* respectively. These values are higher than those measured from the inoculum of the same bacteria, indicating

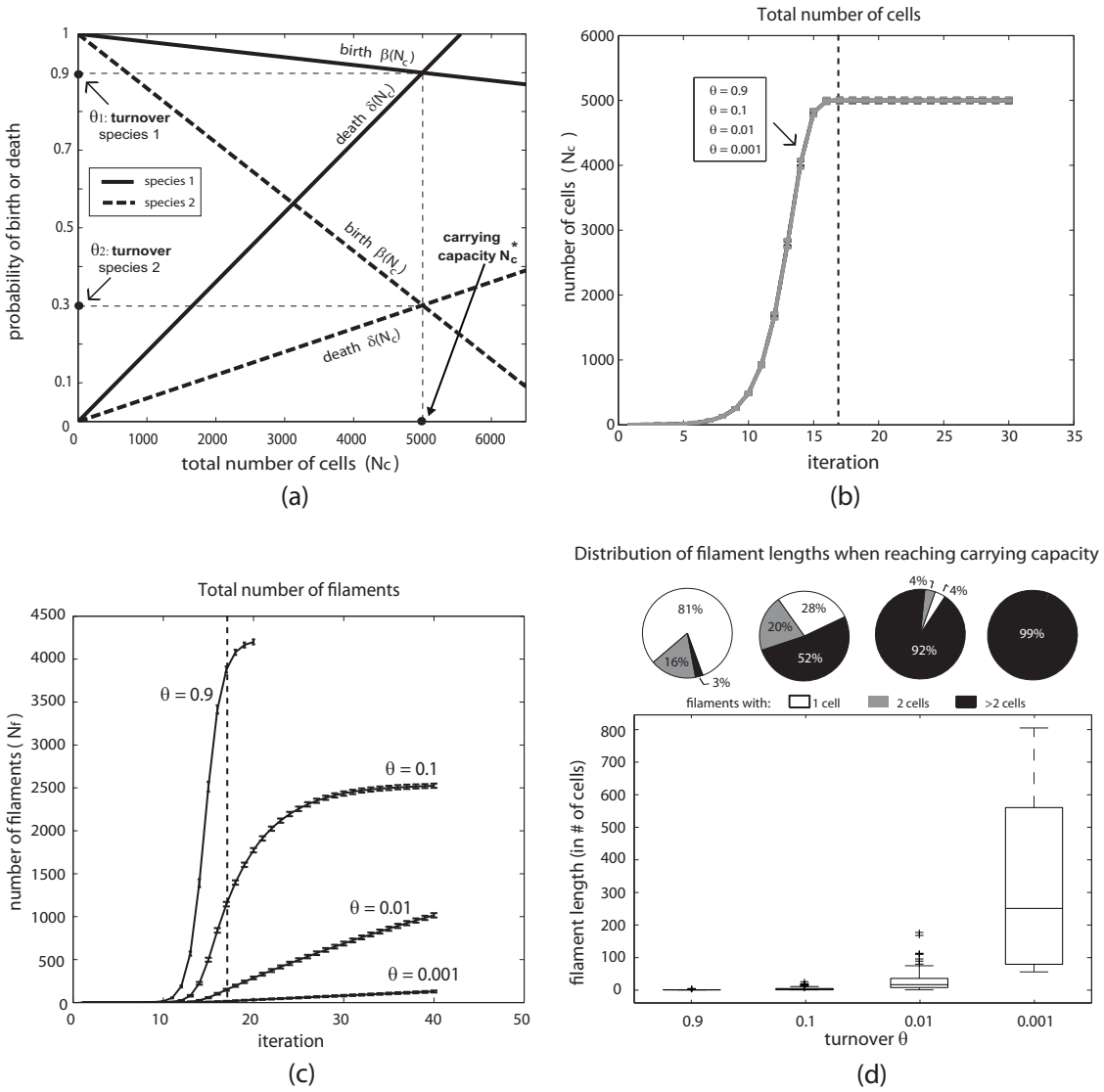


Figure 2.1: Simulation model (a) and results based on 1000 runs (b,c,d). (a) Schematic illustration of birth and death rate functions dependent on total number of cells N_c. The value of the population at which the two rates are the same is defined as the carrying capacity N_c^*. We refer to the value of the rates at the carrying capacity as the turnover rate θ. The plot illustrates that the same carrying capacity can be achieved by species that have a different turnover. (b) For all birth and death functions following the scheme in (a), N_c reaches the steady state following the same growth curve despite the differing turnovers. Dashed line indicates the time at which the carrying capacity is reached (see mathematical definition of N_c^* in § 2.3.1.1). Error bars represent standard deviation of the mean. (c) When the population size reaches the carrying capacity (dashed line), the higher the turnover, the higher the total number of filaments N_f. Nonetheless, after many more iterations, N_f reaches a steady state of about 2500 filaments for all sufficiently small turnovers (plot not shown). Error bars represent standard deviation of number of filaments.(d) Box plots of filament lengths when N_c reaches the carrying capacity. Populations with low turnover have higher median length values than populations with high turnover. The corresponding pie charts on the top show the proportion of single-celled, double-celled and longer filaments in the population.

that cell size of heterotrophs may change over time. For cyanobacteria the mean cell length is less variable, and an approximate cell length of 4.0 μm was used for the conversion.

2.3 Results

2.3.1 Simulation results

2.3.1.1 Turnover and fitness

For the rest of this article, we will refer to the time at which the population reaches its carrying capacity as the first time t such that $N_c(t+1)/N_c(t) \leq 1+\epsilon$, where $N_c(t)$ is the population size at time t and $\epsilon = 1\%$. The growth curve of the total number of cells gives information on the fitness of the strains. Figure 2.1b shows that for any turnover, the total number of cells reaches the carrying capacity after less than 20 generations with the same growth curve (average and standard deviation of 1000 runs is shown). In our case, fitness is represented by net growth rate and does not depend on turnover (see§ 2.2.1). Hence, populations with different turnovers have the same fitness. On the other hand, when the total number of cells reaches the carrying capacity, the total number of filaments (N_f) differs significantly according to the turnover (Figure 2.1c). At this point, N_f ranges from values close to the carrying capacity, corresponding to almost unicellular filaments ($\theta = 0.9$), to values under 20 units ($\theta = 0.001$). Figures 1b and 1c show that the steady states of number of cells and number of filaments are not reached at the same time. This is because the stationary filament length distribution is reached much later than the steady state population density (carrying capacity). In the long run, the number of cells per filament reaches a stable equilibrium of about 2 cells per filament for sufficiently small turnovers, typically less than 0.1 (figure not shown). An analytical support for this result has been obtained by deriving an analogous model in continuous time, as presented in § A.9. For higher turnovers, the effect of discretization leads to a deviation from this equilibrium. Hence, filament length distributions can be affected by turnover rates in temporary phases of the population dynamics, before a stationary distribution is reached. An

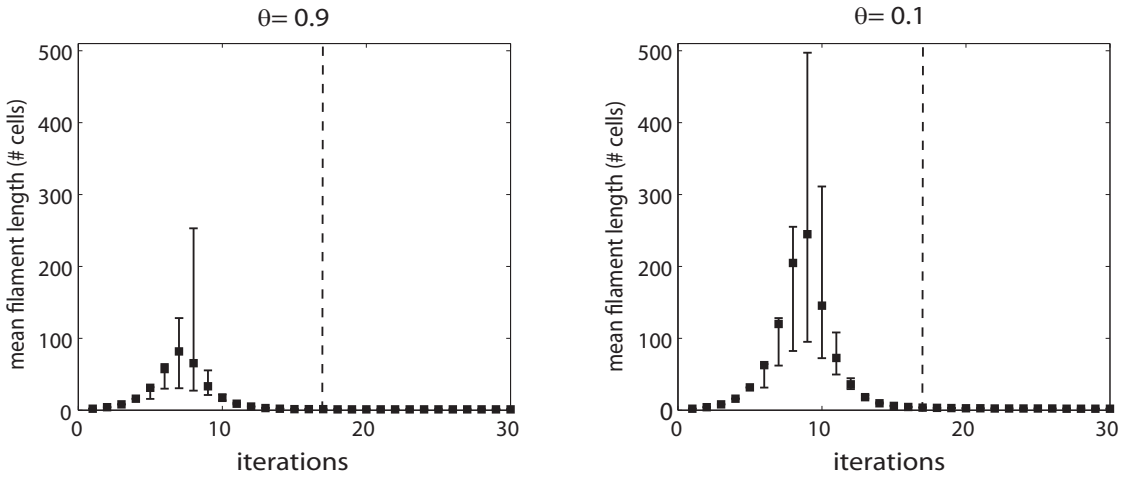


Figure 2.2: Sample simulations showing the mean filament length along successive iterations for two different turnovers. Squares indicate the mean of the mean filament lengths obtained through 1000 runs. Edges of the error bars correspond to the 2.5th and 97.5th percentiles of the distribution of the mean filament lengths in the 1000 runs. In both cases, the mean length reaches a peak in the transient phase, and then decreases gradually toward a state comparable to the one of the beginning. Lower turnovers allow for higher transient peaks. The dotted line indicates the iteration at which the carrying capacity is reached.

example is given by the time when the carrying capacity of the population is reached (17 iterations in this case).

2.3.1.2 Length distribution at the time when carrying capacity is reached

Figure 2.1d shows the box plot of filament lengths when the total number of cells in the population (N_c) has reached the carrying capacity $N_c^* = 5000$ cells. This time point however represents an intermediate stage before a stationary distribution of filament lengths is reached. The corresponding pie-charts indicate the fraction of single cells, double-celled filaments and filaments with more than two cells in the population. When turnover is close to 1 ($\theta = 0.9$), the population consists mainly of unicellular and double-celled filaments. Note that $\theta = 0.9$ means that at carrying capacity, each cell has a 90% chance of dividing as well as a 90% chance of subsequently dying. When $\theta = 0.1$, filaments with one or two cells represent about half of the population. For lower turnovers ($\theta = 0.01, 0.001$), the median values of the length are significantly higher. In the latter cases, filaments with more than

two cells are more than 90%. Figure 2.1d shows that, as the turnover increases, the fraction of short filaments when the population reaches its carrying capacity also increases. The distribution means at this growth stage were statistically compared by means of a one-way ANOVA and a multiple comparison (see § A.6). The means obtained for $\theta = 0.9$ and $\theta = 0.1$ were not significantly different. All the other pairwise comparisons were significant.

2.3.1.3 Length distribution during the transient phase

Figure 2.2 shows two examples where the mean length of the filaments was calculated every iteration. The plot shows the average of 1000 runs, with error bars indicating the 2.5th and 97.5th percentiles for the mean. The dotted line indicates the iteration at which the total number of cells reaches the carrying capacity. An increase to a peak of the mean length followed by a decay towards a steady state is observed in all cases. At the end of the decreasing phase, the mean length is comparable to that in the beginning. The same pattern was observed also for $\theta = 0.01, 0.001$ (see § A.3). An analytical characterization of the filament length dynamics has been carried out in Appendix A (§ A.8), where an approximation for the iteration of first filament breakage has been derived. In order to statistically analyze the average length trend, we considered three reference time points and we compared the distribution of the means at those points using a non parametric hypothesis test. As reference points we chose the starting point, the peak and the carrying capacity. The outcome of the tests provides statistical evidence of a significant increase and subsequent decrease of the average length. Details and results are given in § A.6.

Significant differences are found in the maximum average length achieved. Long filaments, with a mean length up to 250 cells, are observed at a turnover of $\theta = 0.1$. Longer filaments can be achieved with lower turnovers (Fig. A.4). For a higher turnover ($\theta = 0.9$), the average length of filaments at its peak does not exceed 100 cells. The observed trend indicates the prevalence of a cycle in the mean length induced by the birth and death rates, whereby filaments are short in the beginning, longer in the transient phase and again short at the carrying capacity. Moreover, in the transient phase, filaments with low turnover can elongate more

than filaments with high turnover.

2.3.2 Experimental results and model validation

In the theoretical model, the population stays at the carrying capacity once this has been reached. However, since bacterial strains were not cultured in a chemostat, carrying capacities could not be maintained indefinitely. In order to assure that the experimental dataset and the simulation data are comparable, we indicate with a dotted line the time point where the bacterial populations reach their carrying capacity on each plot of the experimental results. This point was determined from the OD curves (not shown). Subsequent to reaching the carrying capacity, the process of filament breakage is eventually accelerated in the empirical case due to a progressive decrease in available nutrients.

In order to prove significant increase and decrease of bacterial filament length, we performed an ANOVA test associated to a multiple comparison on the data based on Tukey's HSD test [204]. Details of the tests for each species and within species are provided in Appendix A (§ A.6).

In the following sections, the time point 0 is represented by the inoculum. The starting culture of heterotrophs was taken from plate colonies and vortexed. Because of nutrient depletion in plate colonies and mechanical breakage through vortexing, the heterotrophic inoculum is composed of mainly single cells. The starting cultures of cyanobacteria were taken from cultures grown for more than one week, gently pipetted and diluted. At inoculum, cyanobacteria filaments had an average filament length of about 45 cells (*N. Muscorum*) and about 75 cells (*A. variabilis*).

2.3.2.1 Mean length trend

Figures 2.3a,b and c show the change in time of the mean length of heterotrophic bacteria *R. lutea*, *F. limi* and *F. aestuarina*. The pattern is similar in all three cases: the mean length increases until a peak, after which it decreases and reaches a value approaching that of the inoculum (see Table A.1 in Appendix A). This trend is recognizable in Figure 2.4, showing micrographs of *F. aestuarina* taken at different time points. Remarkably, experiments A and B on *F. aestuarina* give very

similar results, showing a high repeatability of the observed pattern (Fig. 2.3c). A comparison of the theoretical and *F. aestuarina* length distributions against time is shown in Appendix A (§ A.4, Figure A.6). The mean length of cyanobacteria, tracked for two months, shows a qualitative trend similar to that of heterotrophs, albeit much slower. Both the *A. variabilis* and *N. muscorum* species have a mean-length peak in the transient phase. However, because of the low turnover, they need more time than heterotrophs to reach the stationary distribution, as predicted by the model. When they reach carrying capacity, they are still long.

The pattern observed in heterotrophs is comparable with the simulation results shown in Figure 2.2. The trend of the cyanobacteria, although not so markedly conclusive, gives indication that the model predictions hold also in autotrophs. An additional illustration of the mean length trend of the cultured species is provided in Figure A.5 of Appendix A.

2.3.2.2 Length in the transient phase

Generation times in heterotrophs and cyanobacteria are very different. While the former are observed on the scale of hours, the latter grow on a scale of days/weeks. The difference could be in principle due to the different trophic state of the bacteria. This hypothesis could be tested by means of a comparison between different generation times intraspecifically in the autotrophs or in the heterotrophs. However, this goes beyond the scope of the present work and we proceed relying on the mathematical derivation of section § 2.2.2. According to the relation between generation time and turnover established in § 2.2.2, we assume that the heterotrophs and cyanobacteria have high and low turnovers, respectively. The simulation results presented in § 2.3.1.3 show that the peak value of the mean length in the transient phase increases with decreasing turnover. Figure 2.3 shows that heterotrophic filaments will rarely be longer than 60 cells. The length of cyanobacteria can be significantly higher than that of heterotrophs. *N. muscorum* and *A. variabilis* filaments have up to around 350 and 600 cells respectively. Hence, cultured bacterial populations support the model predictions, indicating that bacteria with a putatively higher turnover (in this case heterotrophic species) are able to elongate less than those with lower turnover (cyanobacteria).

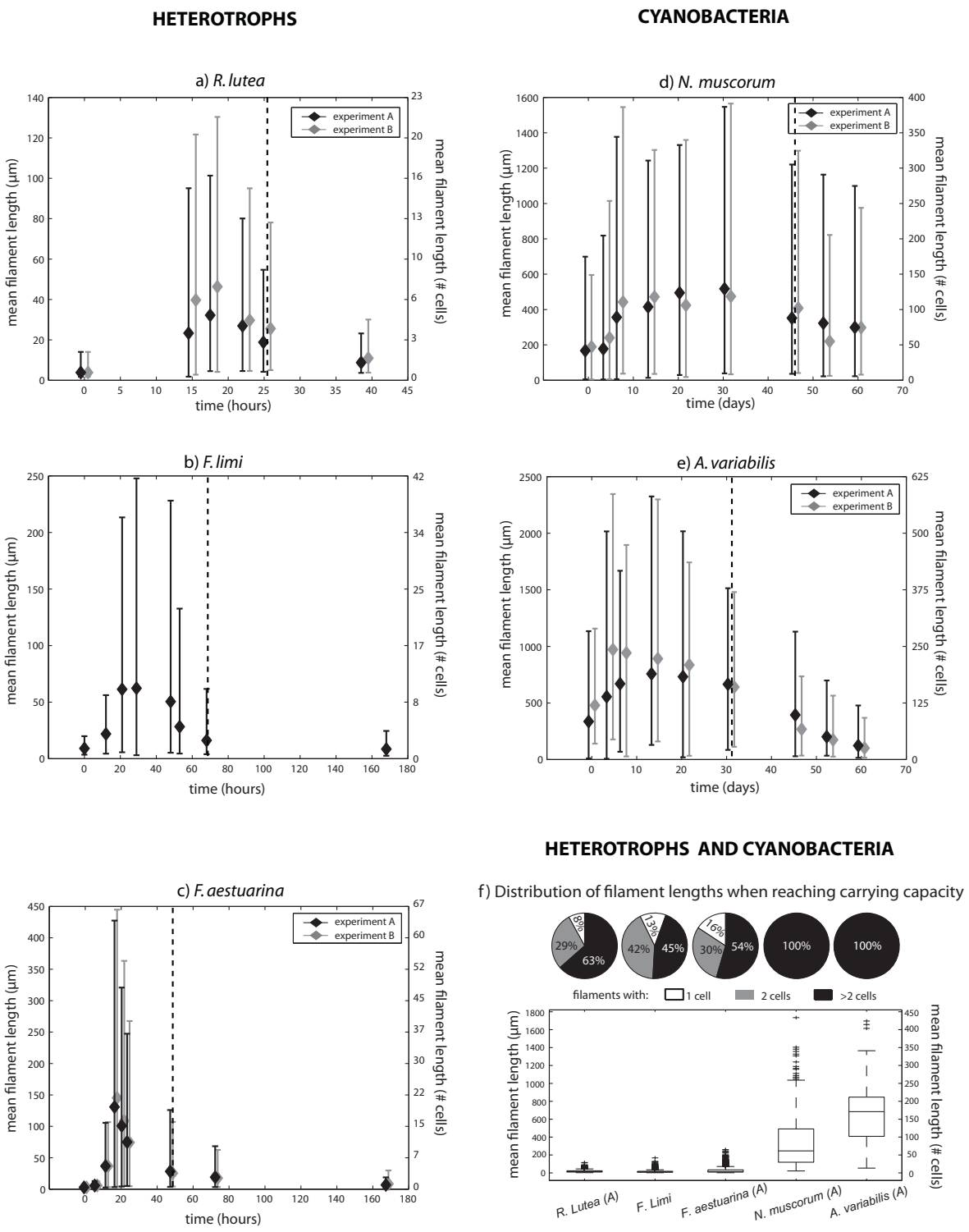


Figure 2.3: (a)-(e) Empirical measurements showing mean length of bacteria against time. Edges of the error bars correspond to the 2.5th and 97.5th percentiles of the measured lengths. Diamonds indicate the mean of measured filament length at each time point. Data from both heterotrophic and photoautotrophic species are shown. The zero time point corresponds to the inoculum. The dotted line indicates the time at which the carrying capacity was reached. (f) Box plots of filament lengths when the bacterial populations reach their carrying capacity. The capital letter next to the species name indicates the experiment used in the plot. The corresponding pie charts on the top indicate the proportion of single-celled, two-celled and longer filaments in the population.

2.3.2.3 Distribution of lengths at the time when carrying capacity is reached

In Figure 2.3f, the box plots and the pie charts show the proportion of short and long filaments in the cultured bacterial populations when they reach their carrying capacity. As for the simulation results, the pie charts illustrate three length classes, namely single cells, double- celled filaments and filaments with more than two cells. The heterotrophs show a considerable amount of short filaments, with a proportion of single cells and double-celled filaments varying across species in the range 8%-16% and 29%-42% respectively. In cyanobacteria, filaments with one or two cells are absent. Figure 2.3f can be compared with Figure 2.1d. Again, the experimental results confirm the theoretical predictions, according to which short filaments at the carrying capacity are more abundant in populations with high turnover.

2.3.2.4 Serial transfer experiment and simulations

If filament elongation is governed by density dependent processes as hypothesized in the model in § 2.2.1, then interventions on such underlying processes should produce testable hypotheses. Experimental manipulation of cell densities is one simple possibility. If the population drops significantly below its carrying capacity, the filaments that survive in this new condition should be able to elongate or at least maintain their original length until the carrying capacity is reached again. We simulated and experimentally carried out a series of successive transfers of bacterial populations, whereby a fraction of the population was transferred to a fresh medium whenever the average length of the original population was close to its peak. Filaments developing in the fresh medium were expected to keep or increase the length reached at the previous stage. Details of the computational and experimental settings are provided in § A.7 of Appendix A. Figure 2.5a shows the average length of filaments obtained by successive transfer simulations. Figure 2.5b shows the average length of *F. aestuarina* tracked along successive transfers. In this case, filament size slightly increases or remains almost constant in most of the cases. The fact that filaments do not elongate so strongly as in the simulations could be due to physical handling (breakage of long filaments while pipetting) or to other culture conditions influencing the elongation that were not considered in the model. The model results indicate that the filaments transferred to the fresh

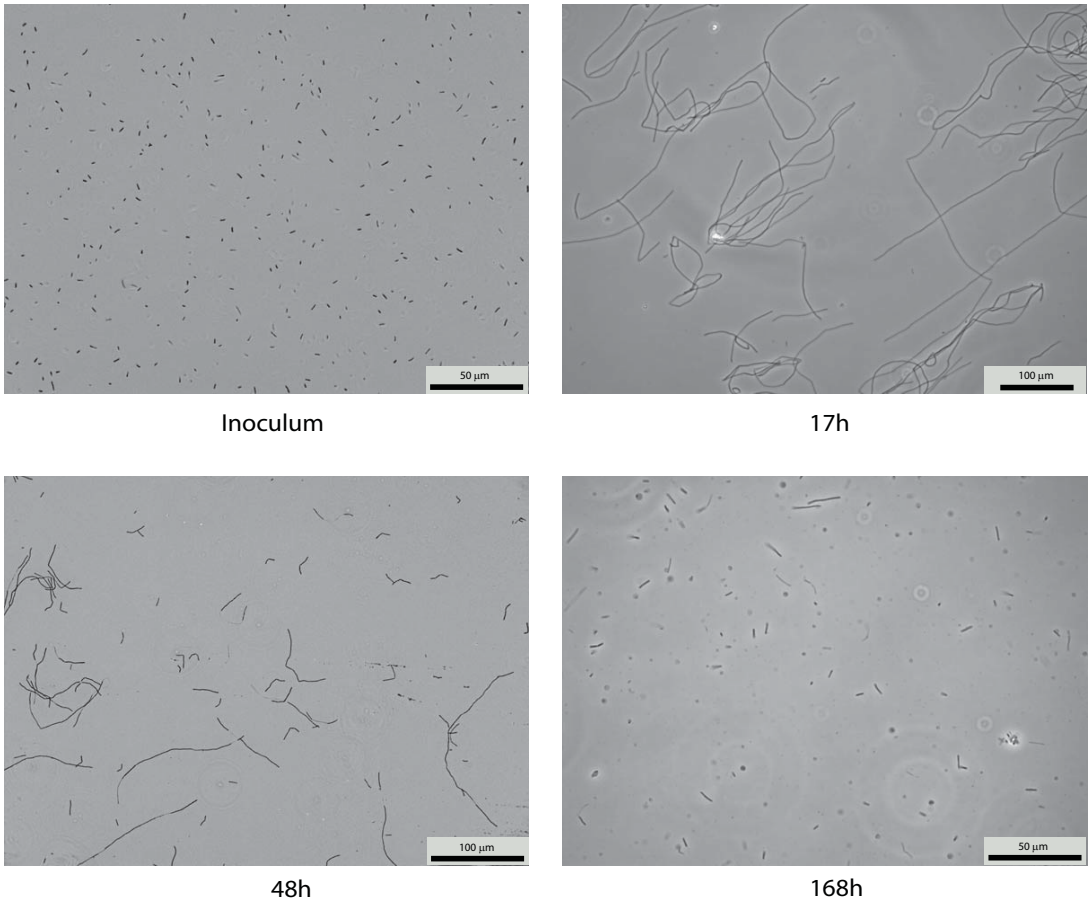


Figure 2.4: Bright field micrographs of *F. aestuarina* bacteria at different stages of their growth. The initial population (inoculum) is composed of many small filaments. After 17 hours, the filaments reach their maximal length. After this peak, the bacteria progressively break towards a mixed population of short and medium filaments as they reach their carrying capacity (48 hours). Eventually, the filament lengths approach the mean of the inoculum.

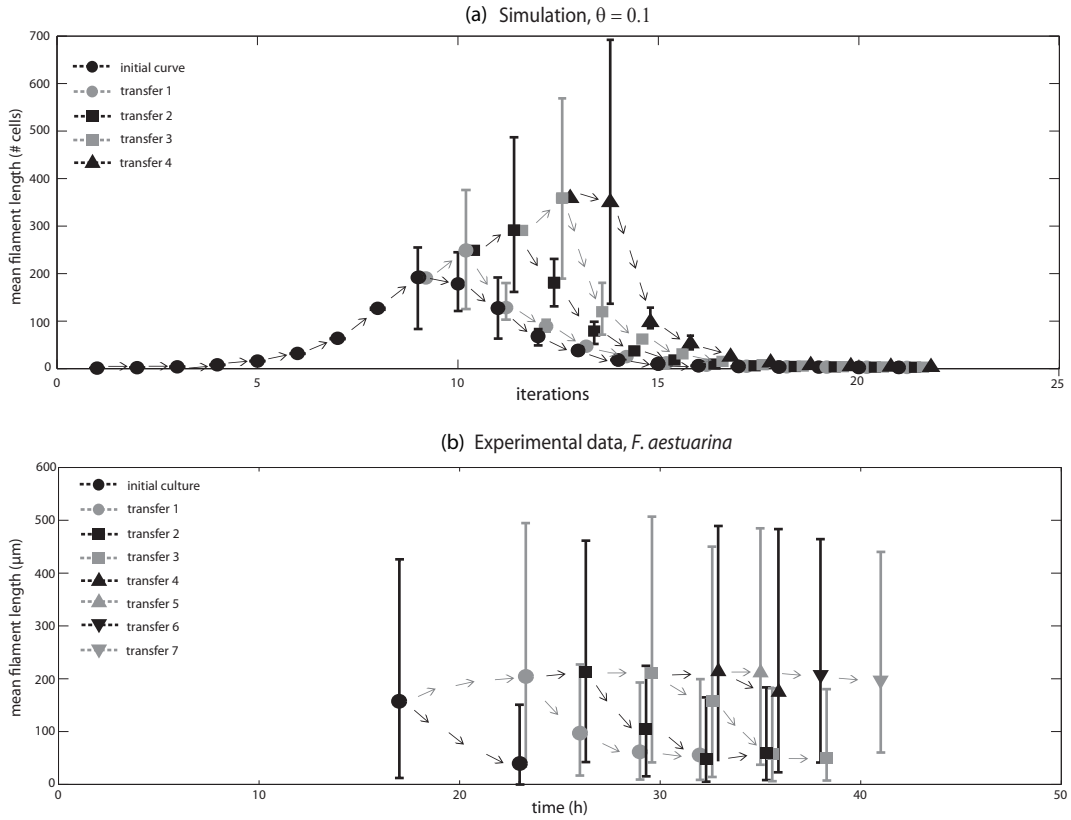


Figure 2.5: Simulations and experiment of serial transfers. Symbol shapes and shading indicate different transfers. Arrows indicate the trend of each culture. (a) Computer simulations of successive transfers. When the average filament length is at its peak, the population density is below the carrying capacity. Whenever a culture reaches this peak, the corresponding average length is used as a starting condition of a new simulation at lower density (transfers 1 to 4). This procedure simulates the transfer of an aliquot of a higher density population to a fresh medium. Edges of the error bars correspond to the 2.5th and 97.5th percentiles of the distribution of the mean filament lengths in the 1000 runs. (b) Average length of *F. aestuarina* during serial transfers to fresh medium. The first transfer was done after 17 hours. For the first five transferred cultures, the mean length was measured and plotted at several successive time points. For subsequent cultures, only one measurement was taken. Edges of the error bars correspond to the 2.5th and 97.5th percentiles of the measured lengths.

medium maintain or increase their length, and eventually, start shortening when they reach their carrying capacity again. The experimental results, although not as strong as the simulations, show that the transferred bacteria are generally able to maintain their length.

2.4 Discussion

Filamentous bacteria can be viewed as a simple example of multicellular individuals. Some of them, such as filamentous cyanobacteria, have been present in the environment since the early stages of life, as indicated by fossil records and phylogenetic analyses [76, 6, 215, 168].The theoretical model shows that at various life stages, two populations with the same fitness and carrying capacity can differ in the composition of filament length classes according to the life history traits regulating their growth. Although the long term stationary distribution of filament length is the same for any sufficiently small turnover rate, the differences due to turnover observed at other growth stages (e.g. transient phase, when reaching carrying capacity) are quantitatively significant and biologically relevant. Filaments belonging to a population with high turnover (and hence short generation time) have moderate lengths at their maximum growth phase. Furthermore, when the population reaches carrying capacity, they are chiefly unicellular. Individuals of a population with low turnover (and hence long generation time) can instead become very long in the transient phase, and still be multicellular when reaching the carrying capacity. The species that we cultured show evidence that the predictions of the model hold. The relation between bacterial generation time and turnover derived in § 2.2.2 helps us to understand the relationship between simulations and experiments. The heterotrophs have a faster generation time than the cyanobacteria (and hence a higher turnover). As a consequence, while the heterotrophic filaments are rarely longer than 60 cells, the cyanobacteria can reach lengths of up to 600 cells. Theory and experiments both show that according to different life history traits, bacteria can cover a wide spectrum of lengths, indicating that multicellularity can be achieved to different degrees. Moreover, a key feature of this model is that the predominance of multicellularity during the life cycle is not necessarily determined by morphology-dependent differences in fitness. Here, the

factors affecting multicellularity are birth and death rates, and hence turnover rate and generation time.

The model indicates a common temporal behaviour for filament length distributions. The mean length increases and peaks during the transient growth phase, and then decreases until it reaches a value approaching the initial conditions. The results from cultured bacterial populations of heterotrophic and photoautotrophic species show that the predicted trend in filament length distribution is robust across species, and is representative of what occurs in nature. The fact that individuals eventually become shorter (and of average length 2 at the stationary distribution) regardless of turnover is due to the effect of cell lysis on a filament. Consider a filament of length L. A cell division (birth) increases its length by one unit, namely the filament length becomes $L+1$. By contrast, after a cell lysis, the expected filament length will be on average $L/2$. The average effect of cell death on length is hence much larger than that of cell birth. This is consistent with the expectation that the average filament length in a population can keep increasing if the death rate remains sufficiently low, as shown in the estimation of the time of first filament split (§ A.8). When the average filament length in the population is at its stationary distribution, cell birth and death have the same effect. This is achievable only if the filaments are of average length 2. This argument is supported by the analysis of the analogous continuous time model derived in Appendix A. In this context, a way to avoid the breakage into small filaments is to decrease the effect of cell lysis on organism size. A solution in this direction is the evolution of a second or third dimension, as in spherical multicellular bacteria [90] or microbial biofilms. In this cases, if we think of a square or a ball of N cells, a cell division will still lead to an individual of $N+1$ cells, but cell death will also only decrease size by one unit, resulting in a size of $N-1$ (as opposed to $N/2$).

There are no previous theoretical and experimental results that are directly comparable to the results presented here. However, the extensive work of [66] and [67] set important precedents and reference points. These works focus on the effects of predators on selection for longer filament lengths. Their results are not directly comparable to those presented here because the death rates due to predators preferentially afflict smaller filaments and single cells. However, they also provide revealing information on the dynamics of filament formation in the

absence of predators. Hahn et al. (1999) show that increasing the nutrient flow and the dilution rate of a chemostat in the absence of predators leads to longer filament lengths. This corresponds to our prediction that in transient growth conditions with lower population densities, filament lengths would get longer. Furthermore, though they do not present extensive data on this, Hahn et al. (1999, p.28) mention that in batch culture, the percentage of multicellular filaments is higher in the exponential phase in comparison to the stationary phase. This situation is again consistent with the predictions presented here.

The model studied here is a general model of filamentous growth, and is not restricted to prokaryotes. Many simple forms of eukaryotic algae grow in a multicellular filamentous form, and the theoretical results presented here are in principle also valid for such eukaryotes. However, it is clear that eukaryotic algae have surpassed their prokaryotic counterparts in their morphological diversity. There are at least six major algal lineages in which filamentous multicellularity has evolved: the Phaeophyta, Rhodophyta, Chrysophyta, Chlorophyta, Charophyta and Embryophyta [129]. However, in each of the latter cases, higher dimensional variants have also evolved, whereupon cell division can occur in more than one planar dimension. This diversity can for example range from the simple globular multicellularity of some of the Volvocales (within the Cholorphyta), to the highly differentiated and developmentally complex vascular land plants (within the Embryophyta). Furthermore, plants can differentiate into diverse developmental modules such as branching nodes, leaves, inflorescences, and root structures. By combining such modules in different combinations, they can access a wider variety of morphologies during their growth process [145, 195]. This flexibility in morphology can translate into both phenotypic plasticity and evolvability.

Developmental life cycles are not unique to multicellular eukaryotes, and an understanding of how they evolved is lacking. Theory and experiments in this article indicate that the size of the individual, in terms of number of cells, can follow a cycle: from unicellular (or short length) to multicellular and back to unicellular (or short length) again. Each time, the initiation of the cycle can be hypothetically achieved by moving cell density below the carrying capacity (e.g. increases in fluid or nutrient availability). The transfer experiment and simulations do not contradict this hypothesis.

In our model, only the birth and death rates play a role in the change in size during the bacterial life cycle. These life history traits are intrinsic properties of every living organism. This pattern, which automatically arises from the interplay between ecology and the filamentous nature of the bacteria, is an emergent property. As such, it can serve as the basis for a primitive life cycle, upon which a more complex developmental program can be subsequently built. As a case in point, some species of *Nostoc* can differentiate into hormogonia during low growth phases characterized by high stress or scarce resources. These hormogonia are short filaments specialized for survival in harsh conditions, which subsequently migrate to a new area before growing into full grown filaments again [118]. Our results suggest that the cycling of filament lengths between growth and static phases of the population is a pre-existing context within which multicellular developmental cycles and differentiation can evolve.

3 The Evolutionary Path to Terminal Differentiation and Division of Labor in Cyanobacteria

Valentina Rossetti, Bettina E. Schirrmeister, Marco Bernasconi and Homayoun C. Bagheri

[*Journal of Theoretical Biology*, 262(1):23-34, 2010]

3.1 Introduction

3.1.1 Multicellularity and the germline-soma divide

Multicellular organisms undergo cellular differentiation in order to perform distinct tasks. A fundamental example is differentiation into germline and somatic cells. This division of labor was first elucidated by Weismann [212, 169] upon studying aquatic animals such as hydrozoans, and green algae of the order Volvocales. He distinguished between germ cells (*Keimzellen*) that contribute cells and hereditary material to the subsequent generation of a multicellular individual, and somatic cells (*Somatische Zellen*) that help in the survival of an individual during its lifetime. In some animals, differentiation into germ cells can be irreversible, referred to as "terminal differentiation." The germline-soma divide is now viewed as a fundamental organizational scheme in complex multicellular organisms, and is central to understanding the interplay between natural selection at the level of the multicellular individual, and competition between its component cells [25, 26].

The separation between a germline and soma is not unique to Eukaryotes, and is also mirrored in differentiated multicellular cyanobacteria [165]. The latter can differentiate into vegetative and heterocystous cells, which are functionally equivalent to germline and soma respectively. Moreover, differentiation into heterocystous cells is terminal. The fact that the same fundamental organizational scheme for the division of labor has independently appeared in such disparate lineages suggests that there may be general conditions that favor the emergence of such an organization. With this view in mind, multicellular cyanobacteria can serve as a model organism for understanding the developmental and ecological conditions that lead to the evolution of terminal differentiation and a germline-soma divide.

Although there is a growing literature on modelling the ecology and population dynamics of nitrogen fixing cyanobacteria [201, 162, 147, 5], the factors that can affect the evolution of multicellularity and differentiation in these organisms has not been examined. In this work we try to approach several fundamental questions. First, we ask what are the fundamental conditions necessary for the evolutionary stability of a terminally differentiated soma in cyanobacteria. Secondly, we ask how differentiation is related to fitness, and how the rate of differentiation can be

optimized in an evolutionary context. Third, we address some of the ecological conditions that may favor the spatial separation of tasks between cyanobacterial cells. Fourth, we examine the phylogenetic history of cyanobacteria in light of our theoretical results.

In the rest of this introduction we discuss the empirical and theoretical background necessary for the models that we subsequently develop.

3.1.2 Multicellularity in cyanobacteria

The cyanobacteria encompass both unicellular and multicellular species, and are among the most ancient multicellular organisms known [171]. Among multicellular species, differentiation into heterocystous forms seems to have a monophyletic origin [205, 177, 202]. Multicellular cyanobacteria such as members of the genera *Anabaena* and *Nostoc* are often present as filaments differentiated into two kinds of cells: vegetatives and heterocysts [220]. Some species also have akinete cells specialized for surviving harsh conditions (hence being similar to spores in their function). We will not deal with akinetes in this study. Vegetative cells are photosynthetic and reproduce by cell division, giving rise to either vegetative or heterocystous cells. They use solar energy and carbon dioxide for the purpose of carbon fixation, and fixed nitrogen in the form of nitrates for building molecules such as amino acids. Fixed nitrogen is produced by heterocysts, whose main task is nitrogen fixation using free atmospheric nitrogen. Heterocysts cannot divide and originate from the division of vegetative cells (a portion of vegetative divisions leads to heterocysts instead of vegetative cells). The need for division of labor between cells that either fix nitrogen or carbon arises from inhibitory chemical interactions between photosynthesis and nitrogen fixation. By having the two chemical reactions occur in different cells, filamentous cyanobacteria can improve the efficiency of nitrogen fixation. In undifferentiated cyanobacteria such as *Synechocystis sp.* or *Oscillatoria sp.*, the main strategy is to have a day and night cycle (cyrcadian rythm) [189, 96, 10, 84, 102], according to which photosynthesis and nitrogen fixation are temporally separated. The interactions among vegetatives and heterocysts can be also framed in the context of cooperation. Heterocysts sacrifice the possibility of reproduction and fix nitrogen for all the cells, in this

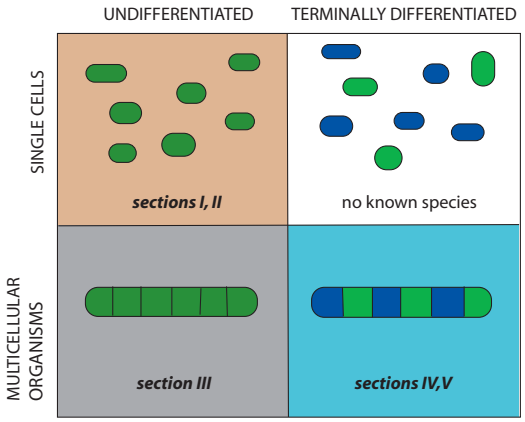


Figure 3.1: Schematic classification of cyanobacterial species based on Rippka et al. (1979).

sense being a fully altruistic entity. Vegetative cells are also cooperative: they do not use all their progeny to pass their genes to the next generation, because part of it will become heterocystous and will lose this ability. If vegetative cells produce few or no heterocysts in order to maximize their reproductive success, they act as defectors.

A detailed classification of the cyanobacteria has been made by Rippka et al. [152]. Cyanobacteria are phenotypically classified into five sections (I-V), which are schematically depicted in Figure 3.1. In the case of heterocystous section IV species such as *Anabaena sp.*, it has been recently established that filaments are truly multicellular, in the sense that the periplasmic space along the filament is continuous [52, 115]. This allows for an exchange conduit for nutrients and other molecules between cells. Given that cyanobacteria are gram negative and possess two membranes, the continuity of the periplasm is acheived via the outer membrane, which forms a unified compartment around a chain of cells, rather than individual cells. Each cell is in turn also encapsulated by its own cytoplasmic membrane. In addition, there is some evidence for direct exchange between the cytoplasms of adjacent cells through membrane channels [126].

3.1.3 Evolution of multicellularity and cooperation

The evolutionary transition between unicellular and multicellular forms involves conflicts between different levels of selection [26, 183, 148, 121]. The benefits associated with multicellularity may for example include size, nutritional advantages, collective protection against antagonists, and division of labor [178, 18, 86]. However, multicellular organization does not automatically imply the existence of differentiation. Undifferentiated multicellularity can have its own advantages over single-celled organization [142, 141, 218].

Once multicellularity has evolved, one can consider the conditions under which cellular differentiation would be advantageous. For example, the division between germline and soma can be analysed as a consequence of the interplay between two fitness components, namely reproduction and survival [212, 123]. Cooperation among cells is fundamental in building a differentiated multicellular organism. Single entities lose the opportunity of selfish reproduction in order to become part of a community of cells. They produce and share nutrients with the others instead of using everything to their advantage, hence increasing the fitness of the multicellular unit [122]. However, such a behavior can be abandoned by defectors (or cheaters), who exploit the cooperative acts but do not contribute to the common good. Following the work of Hamilton [68, 69], various studies have been made about cooperation and selfish behavior using game theoretic approaches [182, 75, 181, 74, 134]. Non cooperative or "cheating" behavior is common in many ecosystems: cheaters can exhibit selective advantages over the competitors [8, 164, 19], but can lead to reciprocal extinction or to stable mutualistic associations [38, 153, 48]. Overexploitation of a common good by cheaters is often referred to as the "tragedy of the commons" [72]. It is known that some kind of subpopulation grouping is required for resolving this problem. The classic explanations are kin selection [68, 69, 184, 55, 105, 214] and reciprocity [203, 8, 107, 70, 105, 196] . Other mechanisms are for example differential dispersal [45, 73], resource supply [20], spatial structuring of the population [132, 133, 47, 127, 142, 141], allowing for the random emergence of association groups [119, 198, 91], or imposing threshold conditions in the rules of the game [9]. Various aspects of these theories have been validated in microbes [24, 194, 207, 148, 63, 114, 214]. For example, assortment

and phenotypic noise can allow the evolution of self destructive-cooperation in *Salmonella thyphimurium* [1], while kin selection limits cheating in the slime mold *Dictyostelium spp.* [24, 61].

Hypercycles, which are autocatalytic networks of enzyme reactions, are another system where the issue of cheating and the importance of population subdivision arises [40, 41, 42]. Hypercycles are susceptible to invasion by "parasitic" enzymes that have reduced catalytic activity for the replication of their target enzyme. It has been suggested several times [40, 42, 185, 43, 119] that one way to escape the problem of parasite invasion in the latter case would be the evolution of compartments or "protocells" that allow different hypercycles to compete. The "stochastic corrector model" of Szathmary and Demeter [198, 183] implements a version of this concept. In a similar vein, an alternative path to achieve population substructuring is the introduction of spatial heterogeneity. [15, 7, 167, 54, 77].

3.2 Methods

At present, there are no known single-celled species of cyanobacteria that terminally differentiate to form collaborative single species consortia as a means to divide labor between nitrogen and carbon fixers (top-right box in Figure 3.1). We model the latter hypothetical scenario (single-celled model) and that of differentiated multicellularity (compartmental model, bottom-right box in Figure 3.1).

3.2.1 Mathematical Models

3.2.1.1 The single-celled model

We consider a single-celled model (Figure 3.2a) where vegetatives, heterocysts and cheater vegetatives compete for nitrate, fixed carbon and solar energy. The vegetative cells convert the solar energy into chemical energy (fixed carbon), while the nitrate is produced by heterocysts. Vegetative cells divide into vegetative and heterocyst cells in different proportions. The cheaters, when present in the system, produce and consume resources at the same rate as the non cheater vegetatives, but they produce less heterocysts —or do not produce them at all. In this model, the resources are shared by all cells living in the environment.

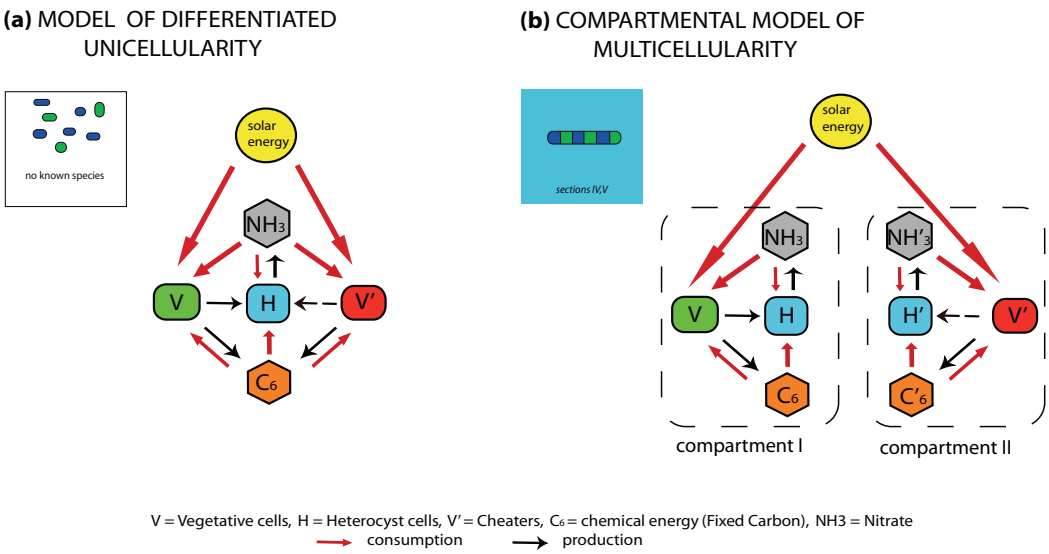


Figure 3.2: Schematic representation of interactions between the variables of the models.

We describe the single-celled model with the following system of ODEs:

$$
\begin{aligned}
\frac{dN}{dt} &= 2a\,H\frac{C}{C+k} - p_3\,N - r(V+V')Z - q\frac{N}{N+k}(V+H+V') \\
\frac{dV}{dt} &= -p_3 V + p_v\,V\,Z \\
\frac{dH}{dt} &= -p_3\,H + p_h\,V\,Z + p_{h'}\,V'\,Z \\
\frac{dV'}{dt} &= -p_3\,V' + p_{v'}\,V'\,Z \\
\frac{dC}{dt} &= c_e Z_L - p_3 C - r(V+V')Z - q\frac{C}{C+k}(V+H+V') - a\,H\frac{C}{C+k}
\end{aligned}
\tag{3.1}
$$

where

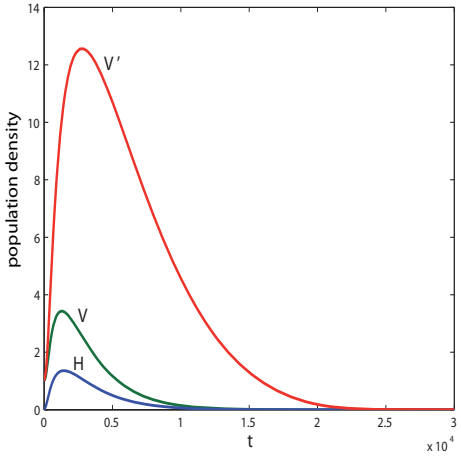
$$\begin{aligned} Z &= Z(N,C) = \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_C\, C} + \frac{1}{k_N\, N} + \frac{1}{k_{NC}\, N\, C}}\,, \\ Z_L &= Z_L(I,G) = \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_I\, I} + \frac{1}{k_G\, G} + \frac{1}{k_{IG}\, I\, G}}\,. \\ I &= \text{irradiance (constant)}, \\ G &= V + V' \text{ (photosynthetic units)}\,, \\ p_v + p_h &= 1,\ \ p_{v'} + p_{h'} = 1 \end{aligned} \tag{3.2}$$

In (3.1), N is the nitrate concentration, V, H and V' are the concentrations of resident vegetative, heterocyst and cheaters cells respectively (cells/unit volume) and C the concentration of chemical energy (sugar in the form of glucose).

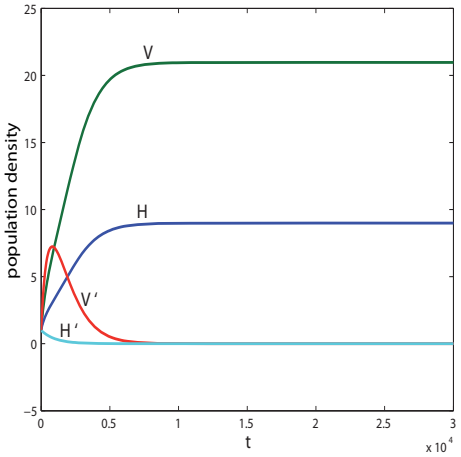
The equations have been built on and can be explained by the following assumptions:
(i) *Reproduction and housekeeping.* The parameter p_v indicates the proportion of vegetative cells originated at any reproduction event. Cheater cells behave similar to vegetative cells, except for their p_v value, designated as $p_{v'}$. Heterocysts are produced in proportion p_h by vegetatives and in proportion $p_{h'}$ by cheaters. Vegetative cells use chemical energy to support dividing ($-rVZ$) and living costs ($-qV\frac{C}{C+k}$). Heterocyst cells use chemical energy for nitrogen fixation ($-aH\frac{C}{C+k}$) and living costs ($-qH\frac{C}{C+k}$). We assume the same death rate p_3 for all cells. The reproduction of vegetatives is regulated by $Z = Z(N,C)$, a Michaelis-Menten type saturation function for two substrates (here nitrogen and sugar).
(ii) *Energetics of nitrogen fixation.* Heterocysts are responsible for nitrogen fixation, which requires about 19 molecules of ATP. Taking into account that a molecule of glucose gives roughly 38 ATP molecules [104], we have

$$\frac{38 \text{ ATP/glucose}}{19 \text{ ATP/fixed N}} \simeq 2 \cdot \text{fixed N/glucose}\,.$$

This relation is the basis for an assumed ratio of 2 molecules of nitrogen produced for every glucose consumed. Nitrogen fixation is only possible whenever carbon (sugar) is available and it is limited by a saturation function dependent on sugar. It is assumed that free nitrogen is not a limiting factor. Decrease in nitrate is due



a Pure cheaters in the single-celled model



b Pure cheaters in the compartmental model

Figure 3.3: Effect of pure cheaters on the resident population. V = vegetatives, H = heterocysts, V' = cheaters, H' = heterocysts in the compartment with cheaters. In (a), all the cell types reach zero, while in (b), cells of the non mutant compartment can grow and reach a positive steady state.

to natural decay ($-p_3 N$), to housekeeping or living cost ($-q\frac{N}{N+k}(V+H)$) and to reproduction of vegetative cells ($-rVZ$).

(iii) *Light harvesting.* The solar irradiance, I, is treated as a static parameter, as is common practice in basic models of photosynthesis [71, 163]. The irradiance is absorbed and transformed into chemical energy (six-carbon-sugar) by vegetative and cheater cells. The total production of sugar depends on I and on the total concentration of photosynthetic units G (PSU). We use the function $Z_L = Z_L(I, G)$ to describe the connection between light harvesting and sugar production. In Z_L, the solar irradiance I is absorbed by the photosynthetic units $G = (V + V')$ present in both the normal and cheater vegetative cells, hence I and G are considered as substrates for Z_L. This type of saturation function ensures that whenever a substrate is limiting the considered cellular activity, the potential increase of the other substrate does not enhance the activity. The solar energy is converted into carbon at a rate c_e. We assume that CO_2 concentrations are not a limiting factor. Carbon is subject to a natural decay ($-p_3 C$).

(iv) *Carbon to Nitrogen consumption ratios.* During exponential growth, the average ratio of carbon to nitrate is roughly C:N$\simeq$6:1 [209] in a bacterial cell. The

uptake of carbon should hence be 6 times higher that that of nitrate. However, considering that a molecule of glucose contains six carbon atoms, we have C:N$\simeq$1:1. The parameter r and q represent the rate of uptake of N and C for reproduction and housekeeping respectively. Based on a 1:1 expectation for C:N content, the coefficients for C:N consumption for both reproduction and housekeeping are assumed to be one.

3.2.1.2 The compartmental model of multicellularity

In the compartmental model (Figure 3.2b), only the solar energy is shared, while different compartments produce and consume their own sugar and nitrate units. The cheater is now an aggregate in which the proportion of vegetative cells produced at each division is higher than in the other. The compartmental model is a simplified representation of multicellularity, in which each compartment represents a distinct multicellular individual. Both compartments compete for the same solar energy source, but each has its own cells, nitrate and chemical energy. The following ODE set describes the dynamics of the compartmental model:

$$\begin{aligned}
\frac{dN}{dt} &= 2a\,H\frac{C}{C+k} - p_3\,N - rVZ - q\frac{N}{N+k}(V+H)\\
\frac{dV}{dt} &= -p_3 V + p_v\,V\,Z\\
\frac{dH}{dt} &= -p_3\,H + p_h\,V\,Z\\
\frac{dC}{dt} &= \frac{V}{V+V'}c_e Z_L - p_3 C - rVZ - a\,H\frac{C}{C+k} - q\frac{C}{C+k}(V+H)\\
\frac{dN'}{dt} &= 2a\,H'\frac{C'}{C'+k} - p_3\,N' - rV'Z' - q\frac{N'}{N'+k}(V'+H')\\
\frac{dV'}{dt} &= -p_3 V' + p_{v'}\,V'\,Z'\\
\frac{dH'}{dt} &= -p_3\,H' + p_{h'}\,V'\,Z'\\
\frac{dC'}{dt} &= \frac{V'}{V+V'}c_e Z_L - p_3 C' - rV'Z' - a\,H'\frac{C'}{C'+k} - q\frac{C'}{C'+k}(V'+H')
\end{aligned} \tag{3.3}$$

where

$$
\begin{aligned}
Z = Z(N,C) &= \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_C\, C} + \frac{1}{k_N\, N} + \frac{1}{k_{NC}\, N\, C}}\,, \\
Z' = Z'(N,C) &= \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_C\, C'} + \frac{1}{k_N\, N'} + \frac{1}{k_{NC}\, N'\, C'}}\,, \\
Z_L = Z_L(I,G) &= \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_I\, I} + \frac{1}{k_G\, G} + \frac{1}{k_{IG}\, I\, G}}\,. \\
I &= \text{irradiance (constant)}, \\
G &= V + V' \text{ (photosynthetic units)}\,, \\
p_v + p_h &= 1, \;\; p_{v'} + p_{h'} = 1
\end{aligned}
\tag{3.4}
$$

Variables N, V, H, C respectively represent nitrate, vegetatives, heterocysts and chemical energy concentrations of the first filament respectively, while N', V', H', C' represent the corresponding variables for the second compartment. The functions Z and Z' have the same meaning as in the single-celled model, except that in this case, they are functions of the respective nitrate and chemical energy of the two compartments. Competition for light between compartments is expressed in the function Z_L, by the variable $G = V + V'$. Light harvesting is due to both compartments, but the terms $\frac{V}{V+V'}$ and $\frac{V'}{V+V'}$ indicate that the income of sugar into the different aggregates is mediated by the concentration of photosynthetic units belonging to the corresponding compartments, hence allowing competition for light. The same assumptions (i)-(v) for system (3.1) listed in section 3.2.1.1 hold also for system (3.3).

3.2.2 Numerical analysis of the models

Due to the high nonlinearity of the equations, we do not derive the analytical expression of the steady states of the system nor do we analytically carry out stability analysis. However, numerical simulations indicate that both systems can evolve towards three different kinds of steady state: one corresponding to the extinction of resident and cheater populations (Y_1), one in which only the resident population survives (Y_2), and one in which the cheaters overcome the resident population (Y_3). Parameters listed in Table 4.1 have been used in the

Parameter description	*Symbol*	*Value*	*Unit*
Uptake of N and C for reproduction	r	5	$mol \cdot cells^{-1}$
Uptake of N and C for housekeeping	q	0.8	$mol \cdot cells^{-1} \cdot s^{-1}$
Uptake of C for N-fixation	a	1	$mol \cdot cells^{-1} \cdot s^{-1}$
Decay rate	p_3	0.001	s^{-1}
Irradiance	I	1000	$\mu E \cdot cm^{-2} \cdot s^{-1}$
Rate of energy conversion	c_e	0.8	$mol \cdot cm^{-3}$
Total stoichiometric concentration	r_0	0.1	
First order rate constant	k_0	10	s^{-1}
Nitrate specificity constant	k_N	10	$mol^{-1} \cdot cm^3 \cdot s^{-1}$
Carbon specificity constant	k_C	10	$mol^{-1} \cdot cm^3 \cdot s^{-1}$
N-C product specificity constant	k_{NC}	1	$(mol^{-1} \cdot cm^3)^2 \cdot s^{-1}$
Irradiance specificity constant	k_I	10	$\mu E^{-1} \cdot cm^2$
PSU specificity constant	k_G, k_V	10	$cells^{-1} \cdot cm^3 \cdot s^{-1}$
Irradiance-PSU product specificity constant	k_{IG}, k_{IV}	1	$cells^{-1} \cdot \mu E^{-1} \cdot cm^5$
Transformed Hill sine/cosine functions	γ	30	

Table 3.1: Parameters values used in the simulations. Abbreviations: E, Einstein; PSU, photosynthetic units.

simulations as default parameters. Structural stability of the models has been tested by random sampling of parameters and initial conditions in $\mathbb{R}^{21}$ and in $\mathbb{R}^{22}$ for the single-celled and compartmental model respectively (results in Appendix B). Numerical integration has been performed using a variable order solver based on linear implicit multistep methods, implemented in function `ode15s` of Matlab (http://www.mathworks.com/).

3.2.3 Evolutionary stability against cheaters

Using the models in section 3.2.1, we simulate competitions between a resident population and either pure or partial cheaters. Mutation is introduced in the systems in the following ways. In the case of pure cheaters, the latter are considered as the mutant. In the single-celled model, mutants are present in the mixed population from the beginning in a given proportion. In the compartmental model, they are introduced in only one of the compartments, while the other one is preserved. In the case of partial cheaters, in both models and for each mutational event, the

strain with a p_v different from the resident strain is considered as the mutant.

3.2.3.1 Evolutionary optimization of vegetative/heterocyst ratio

We consider competitions between strains that differ in their p_v value, with $0 < p_v < 1$. We simulate consecutive competitions between a resident strain (wild type) and a newly arrived mutant. Each step of the simulation is a mutational event, in which after the competition, the winner strain establishes its p_v value as the wild type for the next generation (see Appendix B for details of the algorithm).

3.2.4 Division of labor in time and space: periodic vs. differentiated cyanobacteria

We model a population of undifferentiated cyanobacteria subject to day/night irradiance cycle by the following ODE system:

$$\begin{aligned}
\frac{dN}{dt} &= 2a\delta_n \frac{C}{C+k} V - p_3 N - rVZ - q\frac{N}{N+k}V \\
\frac{dV}{dt} &= -p_3 V + V Z \\
\frac{dC}{dt} &= c_e Z_L - p_3 C - rVZ - (a\delta_n + q)\frac{C}{C+k}V
\end{aligned} \tag{3.5}$$

where

$$\begin{aligned}
Z = Z(N,C) &= \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_C C} + \frac{1}{k_N N} + \frac{1}{k_{NC} N C}}\,, \\
Z_L = Z_L(I,V) &= \frac{r_0}{\frac{1}{k_0} + \frac{1}{k_I I} + \frac{1}{k_V V} + \frac{1}{k_{IV} I V}}\,.
\end{aligned} \tag{3.6}$$

$$I(t) = A\frac{(\rho(t)+1)^\gamma}{(m^\gamma + (\rho(t)+1)^\gamma)} \tag{3.7}$$

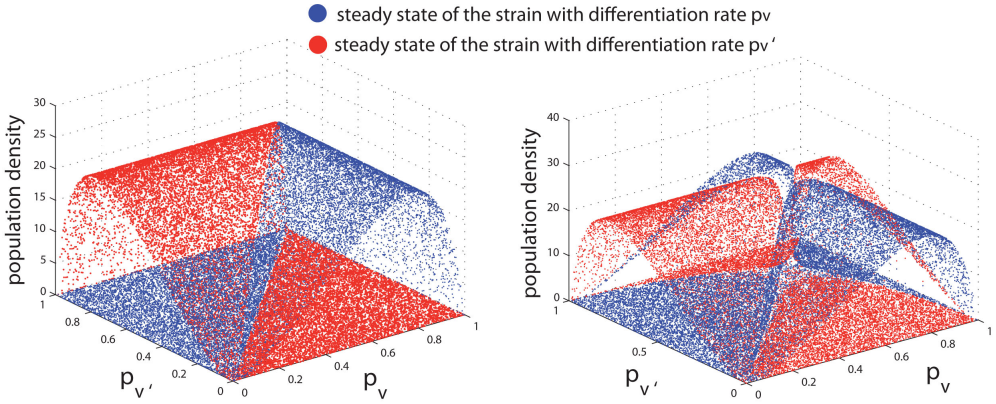
$$\rho(t) = \sin(\frac{\pi t}{12}) \tag{3.8}$$

$$\delta_n = 1 - \frac{(\rho(t)+1)^\gamma}{(m^\gamma + (\rho(t)+1)^\gamma)} \tag{3.9}$$

In (3.5), N, C, and V are fixed nitrogen, carbon and cell concentrations respectively. The irradiance I and the nitrogen fixation function δ_n are Hill transformed sine curves that represent the daylight dependent periodicity [117]. The default parameter values are as in Table 4.1. In this model, we assume that cells do not have a true internal circadian rhythm, but they tune their tasks according to the external day/night alternation. During daylight, the periodic organisms only perform photosynthesis ($I \simeq 1, \delta_n \simeq 0$) because of O_2 inhibition. At night, when the absence of light impede photosynthesis, nitrogen fixation is allowed ($I \simeq 0, \delta_n \simeq 1$). We compare the performance of undifferentiated periodic species in (3.5) with the multicellular differentiated species, when the irradiance is described with a Hill transformed sine curve as in (3.7). To model a population of differentiated cyanobacteria subject to a day/night irradiance cycle, we modify system (3.1) by removing cheaters (V') and using (3.7) for irradiance:

$$\begin{aligned}
\frac{dN}{dt} &= 2a\frac{C}{C+k}H - p_3\, N - rVZ - q\frac{N}{N+k}(V+H) \\
\frac{dV}{dt} &= -p_3 V + p_v V\, Z \\
\frac{dH}{dt} &= -p_3 H + p_h V\, Z \\
\frac{dC}{dt} &= c_e Z_L - p_3 C - rVZ - a\, H\frac{C}{C+k} - q\frac{C}{C+k}(V+H)
\end{aligned} \tag{3.10}$$

where Z, Z_L and I are as in (3.6) and (3.7). As nitrogen fixation is always performed by heterocysts, the function δ_n is not needed. We compare the performance of the two models with different day and night durations, by changing the value of m in I and δ_n. Large and small values of m correspond to long and short dark periods respectively. We map the values of m into a percentage of daylight (details in Appendix B). As the mapping is based on an approximation of the duration of the day, it is not suited to treat neither complete darkness nor absence of darkness. For these cases, we directly set $I = 0, I = 1$ respectively.



(a) Single-celled model:
the strain with the larger pv wins

(b) Compartmental model:
the strain with higher carring capacity wins

Figure 3.4: Steady state density of vegetative cells after competitions between strains differing in their p_v value (plots show representative 15,000 competitions from 115,000). The blue and red dots correspond to strains competing with differentiation rates of p_v (blue) and $p_{v'}$ (red) respectively. (a) In the single-celled model, the strain with the higher p_v wins. Hence, when $p_v > p_{v'}$, the strains with p_v (blue dots) are shown at a higher steady state. When $p_{v'} > p_v$, the red dots are shown at a higher steady state. (b) In the compartmental model, the strain with the higher carrying capacity wins. The ratio of p_v to $p_{v'}$ is no longer the factor that determines which strain wins.

3.2.5 Phylogenetic Analysis of Cyanobacteria

For this study, 16S rRNA gene sequences of 37 cyanobacteria and an outgroup were obtained from GenBank (Table B.1). The ingroup is represented by 9 single celled bacteria from clade I, 4 single celled bacteria from clade II, 14 multicellular bacteria from clade III, 7 multicellular heterocyst forming bacteria from clade IV and 4 branching bacteria from clade V. Our labeling into clades I-V is based on Rippka et al. [152]. *Agrobacterium tumefaciens* was used for outgroup comparison as suggested in previous studies [78, 137]. Details of the analysis are provided in Appendix B.

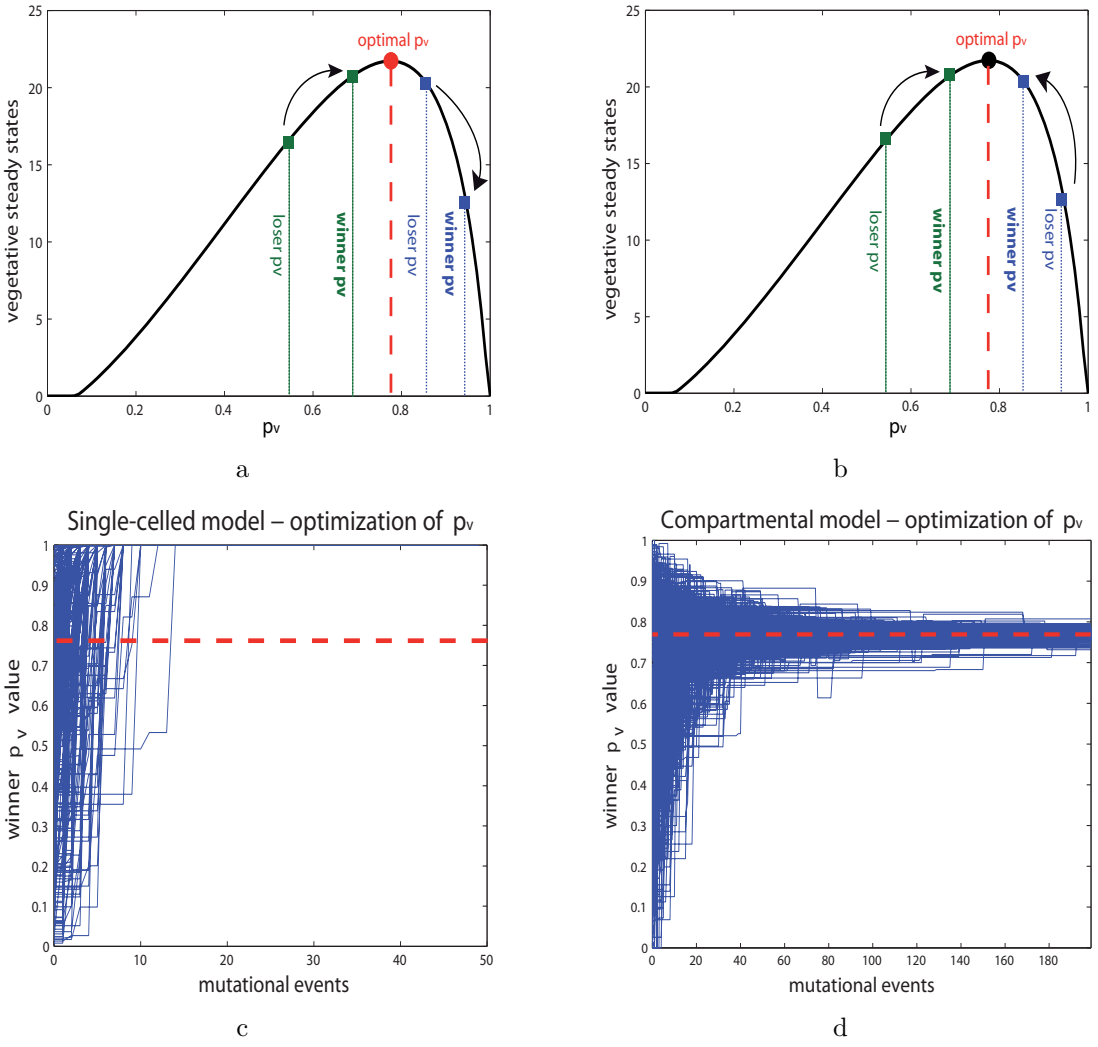


Figure 3.5: Optimization of vegetative to heterocyst cell ratios. Plots (a) and (b) illustrate the winning factor in competitions between strains with different p_v. The parabola-shaped curve represents the carrying capacity as function of p_v. For each case, the carrying capacity of two pairs of strains simulated in isolation are shown as examples (green and blue squares). Next to each strain, we indicate the outcome of the competition within the pair (winner or loser). (a) In the single-celled model, the strain with the higher p_v wins. (b) In the compartmental model, the one reaching the higher carrying capacity wins. Plots (c) and (d) show the outcome of repeated mutational events, at which a competition between resident and mutant strain take place. (c) In the single-celled model, the optimal p_v (red line) is surpassed and $p_v \to 1$. (d) In the compartmental model, p_v tends to the optimal value.

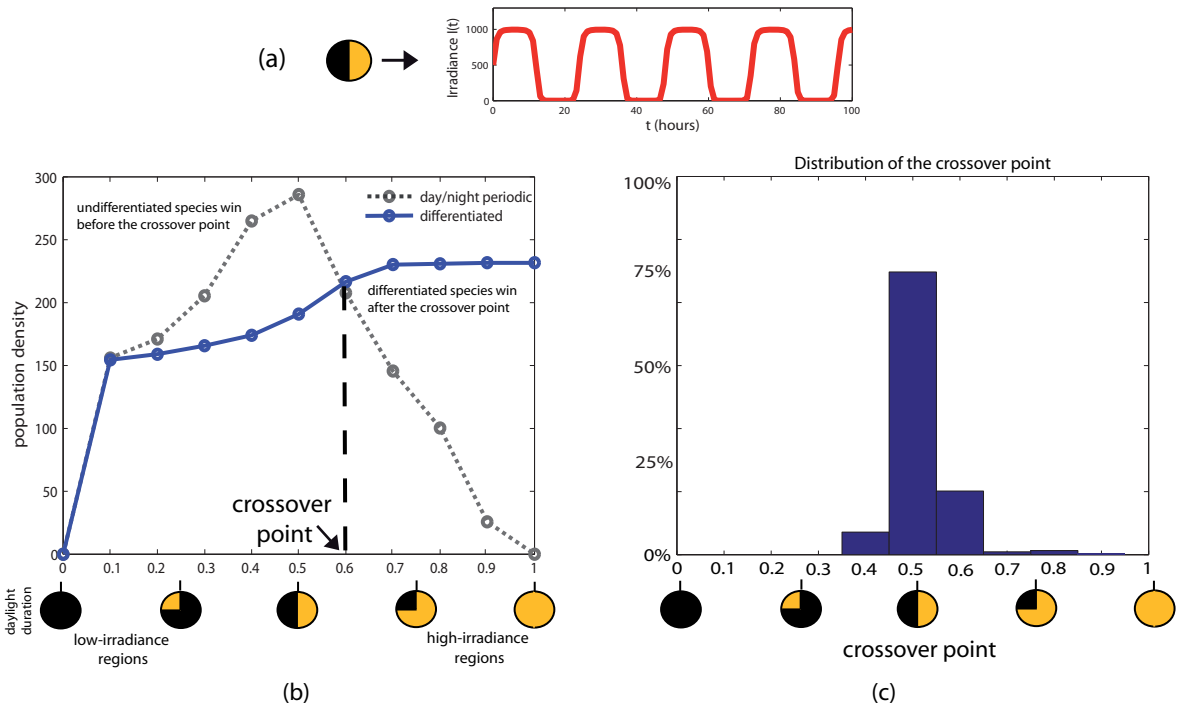


Figure 3.6: (a). The irradiance function $I(t)$ as in equation (3.7), when the duration of day and night are equal. (b). Case study showing the comparison of the steady states of species following the day/night periodicity vs. differentiated species for different external irradiance cycles. Undifferentiated species that follow the day/night alternation reach a higher steady state in environments where the daylight is less or equal than the dark period. The point after which the differentiated species perform better is indicated by the crossover point. (c) Distribution of the crossover points based on 6000 trials. Parameters a, q, r, A were sampled randomly. In 75 % of the cases, the crossover point locates around 0.5.

3.3 Results

3.3.1 Effect of pure cheaters on evolutionary stability

In the single-celled model, introduction of pure cheaters leads to the extinction of the population (Figure 3.3a). Cheaters grow faster and subtract resources from the resident population, which eventually starts decaying after reaching an initial peak. Once the normal vegetative cells are extinct, no entity in the system is able to produce nitrate and the cheaters also die. The collapse of the system in the single-celled model is primarily caused by the fact that the resources are shared

between organisms.

In the compartmental model, a pure cheater cell gives rise to an aggregate of cells that can not sustain itself. The compartmentalization afforded by separate multicellular aggregates (i.e. "multicellular" individuals) allows genetically related cells to protect their resources from a cheater invasion in another aggregate. Hence a cheater can destroy the multicellular aggregate that it arises in, but it cannot destroy the whole population (Figure 3.3b). The basic dynamics of the models without cheaters are provided in Appendix B.

3.3.2 Effect of partial cheaters on evolutionary stability

Partial cheating refers to the situation in which a mutant vegetative cell produces heterocysts in a smaller proportion than the resident vegetative cells. We investigated the criterion that leads to the success of one genotype over the other, when two competing populations differ in their p_v value. We tested the outcomes of competitions in both models using Montecarlo simulations, where pairs of p_v and $p_{v'}$ values were sampled randomly in the interval $[0, 1]$. Vegetative cells belonging to strains with p_v and $p_{v'}$ are indicated by V and V' respectively. Vegetative steady states after competitions are shown in Figure 3.4. Figure 3.4a shows the results for the single-celled model. In this case we find that at a steady state, $V > V'$ when $p_v > p_{v'}$ and similarly, $V' > V$ when $p_{v'} > p_v$. Hence we conclude that in the single-celled model, the winning factor in the competitions is the value of p_v. The strain with the higher p_v outcompete the other. This result holds when randomly sampling through alternative parameter values and initial conditions (see Appendix B).

When tested for the compartmental model, the latter winning criterion does not hold. Figure 3.4b shows that having a higher p_v does not play a role anymore. In the multicellular organization, the winning indicator is the vegetative steady state value that would be reached with the corresponding p_v or $p_{v'}$ value in isolation. This winning indicator is in essence the carrying capacity of each strain when grown separate from the other. Given any two competing strains, we found that the winning strain almost always had the higher carrying capacity when grown in isolation. This in turn depends on the corresponding proportion of vegetative and

heterocyst cells produced during reproduction. Given equal initial conditions for both compartments, the carrying capacity holds true as the winning factor in 97% - 98% of cases when the full parameter space $\mathbb{R}^{22}$ is sampled randomly (details in Appendix B). The 2%-3% exceptions correspond to cases in which (i) a numerical error occurs; (ii) $p_v \simeq p_{v'}$, hence the time required for the populations to stabilize is longer than the simulation time; (iii) The p_v value of the loser is too close to 0, hence the corresponding population can not grow.

3.3.2.1 The optimal rate of differentiation (p_v)

The vegetative cell steady state is dependent on the p_v value. We found that the carrying capacity for the number of vegetative cells is a parabola-shaped function of p_v (Figures 3.5a-b). Hence, a population too rich in vegetative cells would be disadvantaged in comparison to a population with p_v closer to the maximum of the curve. Identical optimality conditions hold in both models. Interestingly, the fact that the optimal p_v values are usually above 0.5 agrees with other theoretical work [218] indicating that the optimal fraction of germline cells in simple multicellular organisms will be higher than that of somatic cells. It also agrees with the high proportion of vegetatives seen in cyanobacteria. Figures 3.5a-b illustrate further what we see in Figures 3.4a-b. Consider the outcome of competitions between a pair of strains with different p_v values. In the single-celled model, the strain with the higher p_v wins, hence the optimal p_v can be surpassed (Figure 3.5a). In the compartmental model, the population with the potential of a higher carrying capacity outcompetes the other, getting closer to the optimal p_v (Figure 3.5b). These results led us to consider the case of repeated competitions, as analyzed in the following section.

3.3.2.2 Evolutionary Optimization of p_v

Stochastic simulations of successive mutational events test the ability of the two strategies to evolve towards the optimal p_v value. Figure 3.5c shows the outcome in the single-celled model. The population always evolves to a full cheater situation, where $p_v = 1$. Hence in this case optimization is not possible. After each mutational event the population with the higher p_v value will go to fixation,

hence increasing p_v towards 1. Figure 3.5d shows the results of the compartmental model. In this case, mutant competitions automatically lead to optimization towards the p_v value that corresponds to the maximum steady state value of vegetatives. Hence, compartmentalization allows populations to evolve towards optimal ratios of vegetative to heterocyst cells.

3.3.3 Duration of daylight and separation of tasks in time and space

We numerically solve systems (3.5) and (3.10) separately. Figure 3.6b shows the steady states of the populations going from permanent darkness to permanent light. In the extreme case of permanent darkness, both species die because of the lack of photosynthesis. In regimes where the percentage of daylight is scarce, the species following the external day/light periodicity reaches a higher carrying capacity than the multicellular differentiated species. The opposite happens when the duration of daylight is much longer than night. In the latter case, as the daylight period is extended, the production of nitrogen decreases in the undifferentiated bacteria until it equals or is less than its consumption. In this situation the steady state population decays to 0. Hence long daylight conveys an advantage to division of tasks in space by means of differentiation. Figure 3.6b shows a crossover point between the steady states, after which periodic species perform worse than the heterocystous. In order to check the occurrence and the location of a crossover point for a more general parameter space, we repeated 6000 times the scan of external daylight percentage for the two models. Each time we sampled randomly the values of parameters a, q, r, A. We recorded the position of the crossover point and we plotted the corresponding distribution through the histogram showed in Figure 3.6c. We can conclude that in the 98,8% of the cases, there was always a crossover point after which the undifferentiated species reach a lower carrying capacity than the differentiated species. In the majority of the cases (75% of the cases), the crossover point is at 0.5, corresponding to a situation where the duration of night equals that of the day. These results indicate that environments where the dark period is significantly longer than the daylight period can be disadvantageous to terminally differentiated species.

3.3.4 Phylogenetic Relationships among Cyanobacteria

Phylogenetic relationships from 16S rDNA sequence of 37 strains of cyanobacteria are shown in Figure 3.7. Cyanobacterial species were grouped into classes I to V as described by Rippka et al. 1979. Therefore only the Bayesian tree is shown. Though the topology is based on Bayesian analysis, character states are colored according to parsimony criteria (such that the least changes occur along the branches). Our analyses confirm the polyphyly of single celled clade I and the multicellular clade III as reported earlier [62, 177]. The multicellular, terminally differentiated clades IV and V, together form a monophyletic group supported by posterior probability (100%) and bootstrap (86%). The latter monophyly has been reported by other studies [205, 59, 202]. Species from the polyphyletic clade III, belonging to the genera *Lynbya*, *Arthrospira*, *Oscillatoria* and *Trichodesmium*, form the sister group of the monophyletic clade IV and V (blue box in figure 3.7). The phylogeny supports the conclusions of the simulations, according to which undifferentiated multicellularity evolved first, and hence made terminal differentiation possible.

3.4 Discussion

At first glance, multicellularity can appear as an obvious prerequisite for cellular differentiation. However, from a logical perspective, alternative developmental strategies are in principle possible. It has been recently emphasized [108] that it is important to strengthen the connection between theoretical models on the evolution of cooperation and explicit empirical cases. The framework we present here is formulated with this goal in mind, whereby we take a mechanistic representation of known biochemical interactions in an important group of organisms (the cyanobacteria), and show how the latter interactions fit into theoretical frameworks that attempt to explain multicellularity and the division of labor.

As discussed in section 3.1.3, practically all solutions for avoiding the tragedy of the commons involve somehow separating the population into competing subsets. The results in sections 3.3.1 and 3.3.2 are no exception to the latter rule. Compartmentalization allows for the protection of vital resources from potential

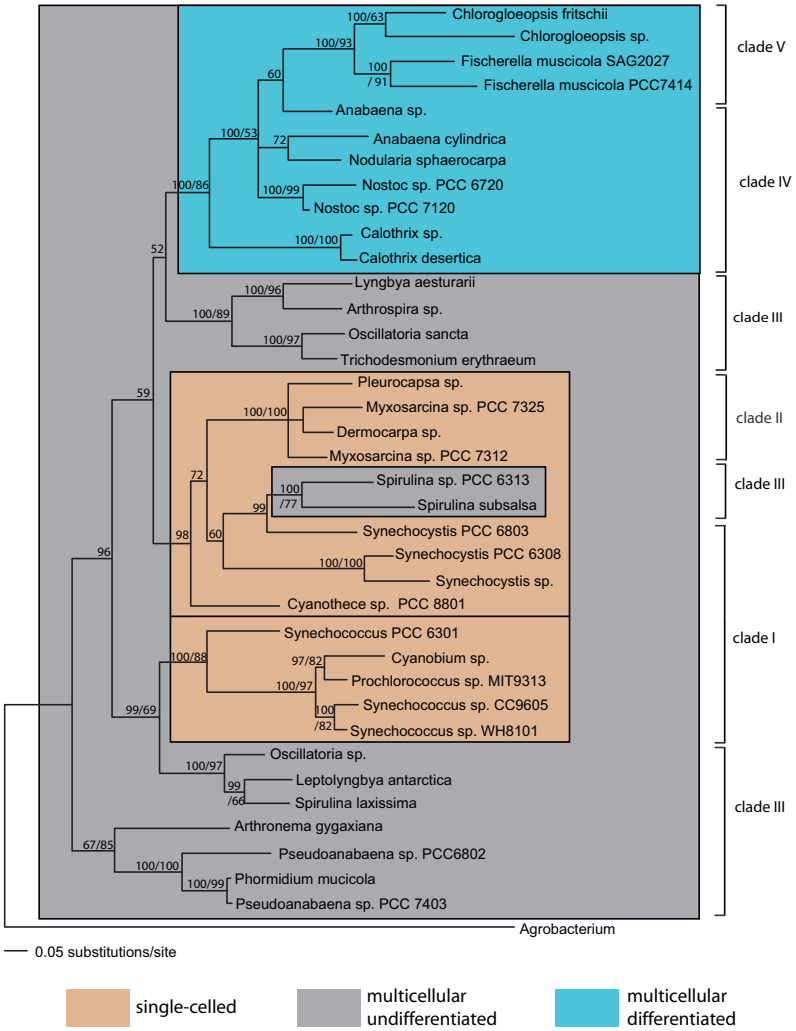


Figure 3.7: Phylogenetic Tree of Cyanobacteria. Bayesian-Tree of 16S rDNA sequences, based on GTR+I+G substitution model, with *Agrobacterium tumefaciens* as an outgroup. Shown at the nodes are only posterior probabilities (>50%) or both posterior probabilities/bootstrap (>50%). Posterior probabilities were calculated from 12,001 trees. Bootstrap values are obtained from 400 pseudo-replicates with maximum likelihood.

disruptive mutations, whose effect can be limited to the compartment they arise in. Furthermore, multicellularity guarantees that cells in a compartment are clones. One could in fact make the argument that the evolution of multicellularity —and thereby compartmentalization— is a mechanistic means by which kin selection becomes "hard-wired" for a population of cooperating cyanobacterial cells. The cell

interactions in the cyanobacterial system also have some structural similarities to a two component hypercycle with a single self-replicating catalyst, the main difference being the lack of hypergeometric growth in the bacterial case. It is noteworthy that conclusions previously derived from the study of hypercycles [42, 119, 198] can also apply to cell interactions and the evolution of multicellularity, indicating the potential generality of the former theory.

The cyanobacterial cell system has also some commonalities with cooperation games. The tragedy of the commons can for example be characterized by games like the Prisoner's Dilemma, where the optimal strategy corresponds to cooperation of both players. However, the cell interactions that we consider do not directly map to a simple n-player game with a payoff matrix. Nonetheless, one may say that populations that converge on the optimal p_v, are in a state where all individuals are cooperating.

As seen in section 3.3.2, the vegetative/heterocyst ratio has an effect on the carrying capacity of the population. Hence the tuning of the proportion of vegetatives upon division (p_v) can lead to the maximization of the carrying capacity. The autonomous optimization of the carrying capacity after repeated mutational events is found to be very different in the two models, as shown in section 3.3.2.2. The single-celled strategy causes the fixation of the variant producing the most vegetative cells, thus converging to the value corresponding to the pure cheater case. Hence, higher fertility in the short term leads to decrease of carrying capacity and eventual extinction of the population in the long term. This explains why this evolutionary step is not observed in nature (see Figure 3.8). On the other hand, the multicellular strategy allows for optimization towards the most profitable proportion of vegetative and heterocyst cells and for the selection and fixation of mutants that correspond to the maximal carrying capacity achievable by the population. Interestingly, this observation agrees with the almost constant vegetative/heterocysts ratio seen in many species of filamentous cyanobacteria [4].

The results from sections 3.3.1 and 3.3.2 exclude the possibility of a transition from the undifferentiated unicellular stage to a differentiated single-celled one. Therefore, to achieve division of labor in cyanobacteria, two other paths are in principle possible: from undifferentiated unicellularity directly to differentiated multicellularity, or via the intermediate step of undifferentiated multicellularity

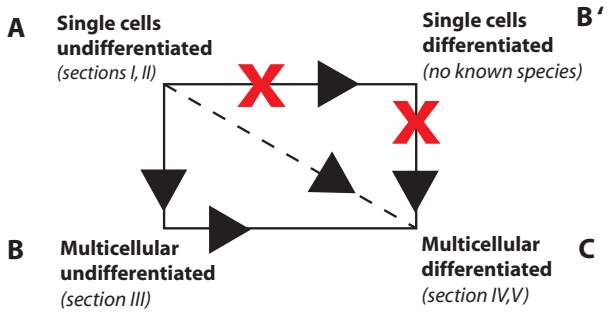


Figure 3.8: The evolutionary paths leading to division of labor in cyanobacteria. A direct transition from A to C is in principle possible. Simulations exclude node B'. The phylogenetic tree supports the path B $\rightarrow$ C.

(Figure 3.8). The outcome of the phylogenetic analyses in section 3.3.4 supports the second alternative, providing empirical evidence that the route to division of labor has included undifferentiated multicellularity. The combination of theoretical and phylogenetic results presented here lead to the conclusion that for the class of interactions occurring in cyanobacteria, multicellularity is a necessary condition for the evolution of terminal differentiation and the optimization of division of labor.

To further understand the ecological factors affecting the evolution of differentiation, we compared the advantages of a spatial separation of tasks over a temporal separation. The geographic distribution of cyanobacteria varies from mild to extreme environments [136]. It is known that environmental factors such as temperature can favor different forms of differentiation in cyanobacteria [188]. However there is at present no clear understanding of the distribution pattern of differentiated vs. undifferentiated cyanobacteria. According to results from our model of bacteria following the day/night periodicity in section 3.3.3, division of labor by means of terminal differentiation is advantageous when the proportion of day is higher than that of night. In the latter case, the undifferentiated cyanobacteria fix nitrogen only in the short dark period. Meanwhile heterocystous cyanobacteria fix nitrogen also during the long day period. Conversely, in cases with scarce daylight, undifferentiated species have an advantage because during the short daylight period all available cells are devoted to light harvesting.

The undifferentiated species that we model simply follow the day/night alter-

nation imposed by the external conditions, without the possibility of development of an internal cycle. The evolution of a self regulated cycle that is independent of external light periodicity can potentially enhance the fitness of such bacteria, because the cycle can then be optimized according to the resource requirements. Hence if the circadian rhythm is optimized, there is the possibility that the undifferentiated circadian species can be also competitive in regions with long daylight periods. A true circadian rhythm has been observed in unicellular cyanobacteria such as *Synechococcus* [124]. Further investigation of the benefits provided by an internal cycle could give an explanation for the maintenance of circadian rhythms in cyanobacteria. On the other hand, the development and regulation of such complex mechanisms is costly for the organism, and one may argue that a high cost of switching could support selection for differentiated species.

Our results on the response to daylight periodicities provide the general conclusion that in an environment with a short light period, selection acts against heterocystous cyanobacteria. In regimes of prevailing darkness, the absence of differentiation and the evolution of a circadian rhythm —or at least the simple adjustment according to the external periodicity— is advantageous. Hence adaptation to long daylight periods can be indicated as a possible reason for the evolution of terminal differentiation in cyanobacteria. The latter is a hypothesis that can be subject to empirical testing. Laboratory experiments can determine the outcome of competitions between undifferentiated and differentiated species under different day/night regimes. In addition, in order to determine seasonal differences between differentiated and undifferentiated species, ecological observations involving sample collection and quantitative measurements of species abundances could be systematically done in different seasons and at different latitudes.

4 Division of labor helps overcome biochemical constraints on biomass production in cyanobacteria

Valentina Rossetti and Homayoun C. Bagheri
[*submitted*]

4.1 Introduction

Cyanobacteria are among the most ancient phyla of photoautotrophic bacteria. They make use of solar light for oxygenic photosynthesis, and many species (diazotrophs) are able to fix nitrogen. A large part of global nitrogen fixation is due to diazotrophic cyanobacteria [27, 35, 125] and cyanobacteria are important oxygen producers [21]. Interestingly, oxygen produced by photosynthesis inhibits the nitrogenase enzyme required for nitrogen fixation. Cyanobacteria evolved two main strategies to overcome this chemical incompatibility. Terminally differentiated cyanobacteria employ a separation in space, whereby nitrogen fixation is confined to somatic cells (heterocysts) endowed with a thick cell wall. Alternatively, undifferentiated cyanobacteria use a circadian rhythm, whereby photosynthesis takes place during the day, and nitrogen fixation during the night. Nitrogen and carbon are dynamically stored in specialized granules [99], or as glycogen, respectively. These cellular structures are subject to adaptation to external conditions [144, 170] and can represent an advantage over algal species [99, 79].

The versatility of Cyanobacteria allow them to be ubiquitous on our planet, although the availability of solar light varies dramatically according to latitude and season. They have been detected in Svalbard [88], Greenland hot springs [155], Baltic Sea [46, 191], North Pacific Ocean [223, 35], Bird Shoal and Bahamas [136], Gulf of Guinea [146], Antarctica [37] and Siberian permafrost [208]. Despite this extensive literature, a clear pattern of cyanobacterial species distribution has not been elucidated nor attempted. Some mathematical models involve a large number of parameters [53, 125], or consider single case studies or distinct habitats. Examples are competition for light and nitrogen between nitrogen fixing and non-nitrogen fixing species [5], comparison of circadian species with different free-running periods [162] or population dynamics and nitrogen fixation rates in the Baltic Sea [190, 46, 112, 188].

In evolutionary history, terminally differentiated cyanobacteria represent an important example of a primitive form of division of labor between germline and somatic cells. This kind of cellular differentiation has been first observed and widely studied in the eukaryotic order of Volvocales [212], where it could evolve along several evolutionary paths [94] as a consequence of a trade-off between reproduction

and survival [123]. More recent modeling approaches, applicable to eukaryotes and prokaryotes, investigated the most advantageous proportion of somatic cells in multicellular organisms [218] and the possibility of rapid evolution of division of labor as a consequence of developmental plasticity [60].

This study aims to elucidate the advantages of spatial and temporal division of labor with regard to biomass production and reproductive success in cyanobacteria. Our model is devoted to investigate the role of geographic location and resource management in cells, and the performance of undifferentiated and differentiated cyanobacteria. Moreover, we compare them with the hypothetical scenario of bacterial organisms that are free from such constraints. The models in this work extend the one developed in [161], by introducing a more detailed representation of light availability and of circadian rhythm.

4.2 Materials and Methods

4.2.1 The mathematical model

We developed a model of cyanobacterial populations that survive through the production and exchange of fixed carbon and nitrogen. We model two cyanobacterial species types and an idealized species used as a hypothetical case, as shown in Figure 4.1. The cyanobacterial species types correspond to undifferentiated (circadian) and terminally differentiated (heterocystous) species. In the former, carbon and nitrogen fixation alternate according to a circadian rhythm. In the latter, they are spatially separated. The species of the idealized model has no biochemical constraints, whereby nitrogen and carbon fixation can be performed in each cell whenever substrates are available. The performance of each species is measured in terms of biomass production and population size. The dependence on geographical location is expressed through the latitude and is included in a function for solar irradiance based on real data of irradiance on Earth's surface. The resource investment of the cell in reproduction and nitrogen fixation occurs at variable rates, represented by parameters r and a respectively. All models consist of a set of ordinary differential equations that were run for a number of iterations corresponding to 40 years. Each species was evaluated at 5 latitudes ranging from 0 to 60 N, for

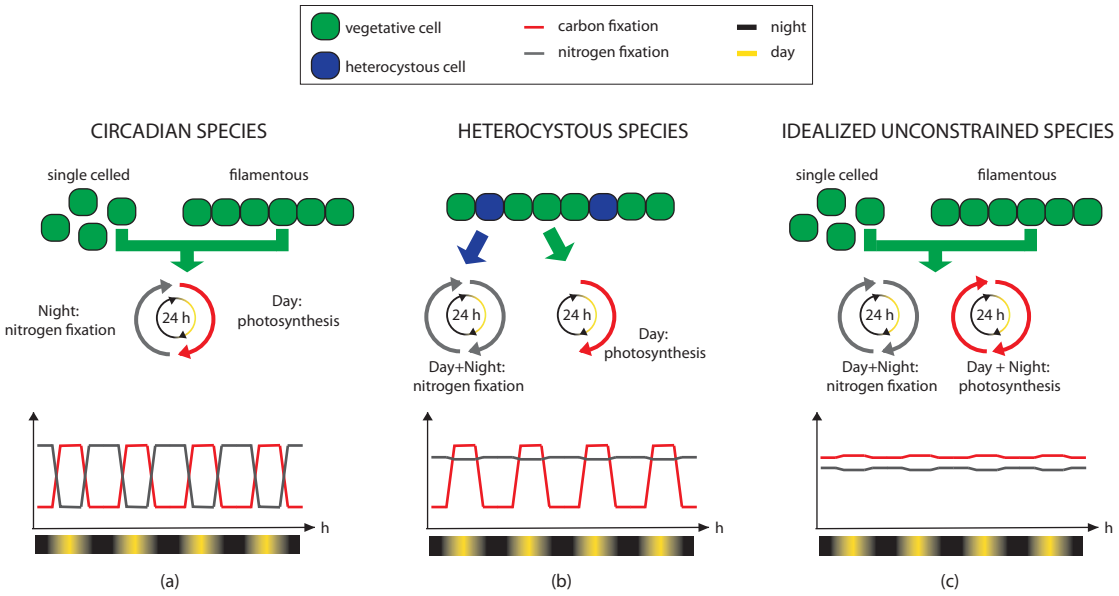


Figure 4.1: Schematic view of the model species. (a) Undifferentiated cyanobacteria (both single-celled and filamentous species) fix carbon and nitrogen according to a circadian rhythm. All cells perform photosynthesis during day and fix nitrogen during night, hence the two processes alternate on a day/night rhythm. (b) In terminally differentiated cyanobacteria, vegetative cells perform photosynthesis during daytime, when light is available. Heterocyst cells provide an anoxic environment where nitrogen fixation can take place at any time, hence limited only by the availability of resources. (c) In the idealized species, nitrogen and carbon fixation can be performed without temporal inhibitory restrictions and are limited only by the availability of substrates. Oscillation amplitude and position of the red and grey lines for carbon and nitrogen fixation are arbitrary and for illustration purposes only.

13 values of parameter r ranging from 10^{-5} to 10 $fmol{\cdot}cells^{-1}{\cdot}h^{-1}$ and for 5 values of parameter a ranging from 0.05 to 5 $fmol \cdot cells^{-1} \cdot h^{-1}$. Assuming a symmetry of environmental conditions between northern and southern hemispheres, the model takes into account only northern latitudes. In the following, we present the systems of equations representing each species type. Concentrations of vegetative and heterocystous cells are indicated with V and H respectively. Concentrations of nitrogen and carbon are represented by N and C respectively. Default values and units of model parameters are listed in Table 4.1.

4.2.2 Circadian species.

This model represents undifferentiated cyanobacteria with a circadian program. It does not distinguish between single-celled and filamentous species, as it considers both as a set of circadian-regulated cells producing and sharing carbon and nitrogen. Nitrogen fixation and photosynthesis are regulated by an internal clock that follows daily light availability. The endogenous rhythm of the cellular activities is expressed by the functions δ_c and δ_n: these functions are multiplied by the carbon and nitrogen fixation terms and inhibit these processes during night and day, respectively. The system of equations is given by

$$\begin{aligned} \frac{dN}{dt} &= 2a\delta_n \frac{C}{C+k} V - rVZ \\ \frac{dV}{dt} &= -p_3 V + r_d V Z \\ \frac{dC}{dt} &= r_p \delta_c Z_L - rVZ - a\delta_n \frac{C}{C+k} V \end{aligned} \tag{4.1}$$

where

$$\begin{aligned} Z = Z(N,C) &= \frac{r_0\, N\, C}{k_0 + \frac{1}{k_C\, C} + \frac{1}{k_N\, N} + \frac{1}{k_{NC}\, N\, C}}\,, \\ Z_L = Z_L(I,V) &= \frac{r_0}{k_0 + \frac{1}{k_I\, I} + \frac{1}{k_V\, V} + \frac{1}{k_{IV}\, I\, V}}\,. \end{aligned} \tag{4.2}$$

4.2.3 Heterocystous species.

In this model we consider terminally differentiated species. These possess heterocystous cells, where nitrogenase is protected from oxygen. Hence the nitrogen production term does not have any inhibition coefficient δ_n, however only heterocysts can fix nitrogen. Carbon fixation follows the external light/dark alternation. Vegetative cells differentiate into heterocysts with a rate $p_h = 1 - p_v$. The default value $p_v = 0.85$ was chosen as an approximation of the values observed in nature. The set of equations is given by

$$\begin{aligned}
\frac{dN}{dt} &= 2a\frac{C}{C+k}H - rVZ \\
\frac{dV}{dt} &= -p_3 V + r_d p_v V Z \\
\frac{dH}{dt} &= -p_3 H + r_d p_h V Z \\
\frac{dC}{dt} &= r_p Z_L - rVZ - a\,H\frac{C}{C+k}
\end{aligned} \tag{4.3}$$

where Z and Z_L are as in (4.2).

4.2.4 Idealized unconstrained species.

This hypothetical species type represents an ideal cyanobacterial species in which biochemical constraints due to the inhibition of oxygen on nitrogenase are absent. This implies that nitrogen and carbon fixation can take place in each cell and at any time, as soon as the necessary resources are available. No such species is known. The system of equations describing this model is similar to that of the circadian species, however nitrogen and carbon fixation are not regulated by the

δ_n and δ_c functions anymore:

$$\begin{aligned}
\frac{dN}{dt} &= 2a\frac{C}{C+k}V - rVZ \\
\frac{dV}{dt} &= -p_3V + r_dV\,Z \\
\frac{dC}{dt} &= r_pZ_L - rVZ - a\frac{C}{C+k}V
\end{aligned} \tag{4.4}$$

where Z and Z_L are as in (4.2).

4.2.5 Model explanation.

In systems (4.1), (4.3) and (4.4), the parameter r is the rate at which nitrogen and carbon are invested per cell division. Carbon is consumed at rate a to be invested in nitrogen fixation. Nitrogen is fixed at a rate $2a$. The reproduction term is represented by the function Z, whose substrates are carbon and nitrogen. Carbon fixation is expressed by the function Z_L that depends on the irradiance I and on the vegetative cells (considered as photosynthetic units). Since solar light only depends on time, but is not an explicit variable of the system, the use of photosynthetic units as limiting factor prevents the exponential growth of the population. The reader can refer to [161] for a more detailed description of the equation structure.

Solar light is represented by I, an antilogistic function [117] given by

$$I(t) = A\frac{e^{\beta(\gamma-\alpha)}}{1+e^{\beta(\gamma-\alpha)}}\,. \tag{4.5}$$

where

$$\begin{aligned}
&\gamma = \cos((t-\phi)\frac{\pi}{12})\,, \quad \alpha = -\frac{-\beta\,\rho + \log(-\frac{x_s}{(-1+x_s)})}{\beta}\,, \\
&\rho = \cos((s-\phi)\frac{\pi}{12}) \quad \beta = 70\,, \quad \phi = 12\,, \quad x_s = 0.00001\,, \\
&A = \text{max irradiance}\,, \quad t = \text{time}\,, \quad s = \text{sunrise time}\,.
\end{aligned} \tag{4.6}$$

The value of maximum irradiance A depends on the month and on the latitude and it based on real data (courtesy of the Swiss Federal Research Institute WSL,

Parameter description	*Symbol*	*Value*	*Unit*
Resource investment in reproduction	r	10^{-5} to 10	$fmol \cdot cell^{-1} \cdot h^{-1}$
Resource investment in N-fixation	a	0.05 to 5	$fmol \cdot cell^{-1} \cdot h^{-1}$
Decay rate	p_3	0.001	h^{-1}
Max. Irradiance	A	1000	$\mu E \cdot m^{-2} \cdot h^{-1}$
Half saturation constant	k	100	$fmol \cdot \mu m^{-3}$
Total stoichiometric concentration	r_0	1	-
First order rate constant	k_0	1	-
Nitrate specificity constant	k_N	10	$fmol^{-1} \cdot \mu m^3 \cdot h^{-1}$
Carbon specificity constant	k_C	10	$fmol^{-1} \cdot \mu m^3 \cdot h^{-1}$
N-C product specificity constant	k_{NC}	1	$(fmol^{-1} \cdot \mu m^3)^2 \cdot h^{-1}$
Irradiance specificity constant	k_I	10	$\mu E^{-1} \cdot m^2$
PSU specificity constant	k_G, k_V	10	$cells^{-1} \cdot \mu m^3 \cdot h^{-1}$
Irradiance-PSU product specificity constant	k_{IG}, k_{IV}	1	$cells^{-1} \cdot \mu E^{-1} \cdot \mu m^5$
Transformed Hill sine/cosine functions	γ	30	-
Photosynthesis rate	r_p	6	$fmol \cdot \mu m^{-3} \cdot h^{-1}$
Cell reproduction rate	r_d	0.03	h^{-1}

Table 4.1: Parameters values used in the simulations. For parameters r and a ranges are indicated. Abbreviations: E, Einstein; PSU, photosynthetic units.

Birmensdorf). The sunrise time is determined according to day and latitude (timeanddate.com/worldclock/sunrise.html). Plots of the irradiance function and details regarding the approximation of real data with the function I can be found in Appendix C.

Circadian species are characterized by an endogenous clock that regulates the timing of carbon and nitrogen fixation. Similar to the irradiance case, we use antilogistic functions to model carbon and nitrogen fixation rhythm, expressed by δ_c and δ_n respectively:

$$\delta_c = \frac{e^{\beta(\gamma-\alpha)}}{1+e^{\beta(\gamma-\alpha)}}, \quad \delta_n = 1-\delta_c\,, \tag{4.7}$$

where γ, α and ρ are as in (4.6).

4.2.6 Performance measures.

All differential equations were numerically integrated with Matlab. Each species was simulated separately in order to determine its performance without inter-species competition. The performance of the populations was measured in two ways: first as total population size (vegetative cells for circadian and unconstrained species, vegetative plus heterocysts for the differentiated species); second, as biomass accumulation. The proxy for biomass in this case is the total amount of carbon given by

$$\text{biomass} = \frac{r}{r_d} \cdot V \,, \tag{4.8}$$

where V is the population size expressed as number of cells per unit volume. In (4.8), the ratio r/r_d between the cell investment of carbon for reproduction and the cell division rate represents the amount of carbon produced per cell. By multiplying this ratio for the total number of cells, we obtain the total amount of carbon (measured in fmol/μm^3).

4.3 Results

4.3.1 Population dynamics.

The model developed for this work is entirely deterministic, though the analytical complexity of the equations allows only for a numerical study of their dynamics. Figure 4.2 shows the daily population size over 40 years for sample single runs. The system does not reach a steady state, rather it presents a dynamic characterized by oscillations and peaks similar to that of chaotic systems. This can be explained by the fact that although the pattern of light availability is repeated identically every year, the population size at the beginning of each yearly cycle differs from the one of the previous year. Moreover, the speed at which light availability changes is higher than the speed at which the cyanobacteria population can equilibrate at the new external conditions. Hence a steady state cannot be reached.

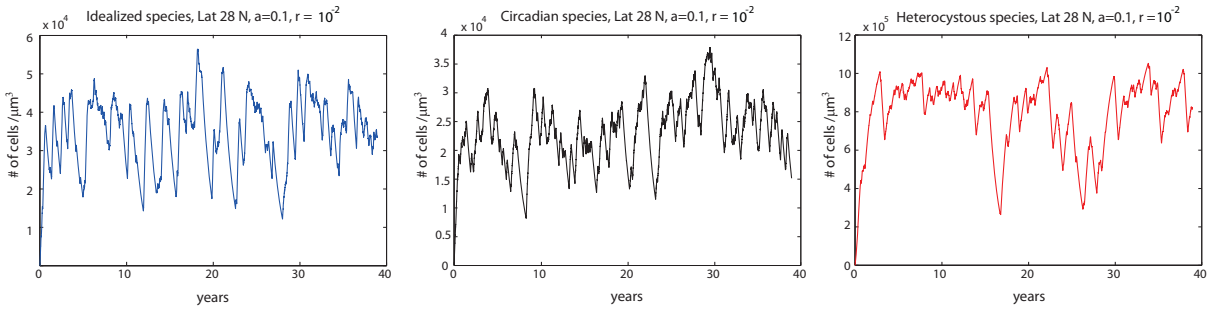


Figure 4.2: Sample runs of the three modeled species simulated for 40 years.

4.3.2 Effect of latitude and cell investment on performance.

In Figure 4.3, box plots of biomass and number of cells over a 35 years period are shown at different values of r and latitude. r can be interpreted as the rate of nitrogen and carbon investment into population growth per cell division. For each value of r and latitude, data are plotted for the value of a that provides the highest mean biomass. a can be interpreted as the rate of carbon investment in nitrogen fixation. Up to $r = 5 \cdot 10^{-1}$, the population biomass of all three species generally increases with increasing r. For $r > 5 \cdot 10^{-1}$, biomass flattens or even decreases. This trend is less marked at latitude 60 N. While the biomass of heterocystous and unconstrained model species ranges between comparable values, that of circadian species is clearly lower than the others.

The trend of the number of cells is very different from that of biomass. For all species, the number of cells decreases rapidly with increasing values of r. Circadian species exhibit always lower values than the other two species types. For the smallest value of r, heterocystous species reach higher population size than the unconstrained species.

4.3.3 Performance comparison at optimal cell investment.

For each latitude and species type, there is a pair of a and r values that provide a maximum average biomass. We refer to these values as optimal (a, r) pairs. In Figure 4.4a, we compare the biomass distribution of the two species obtained

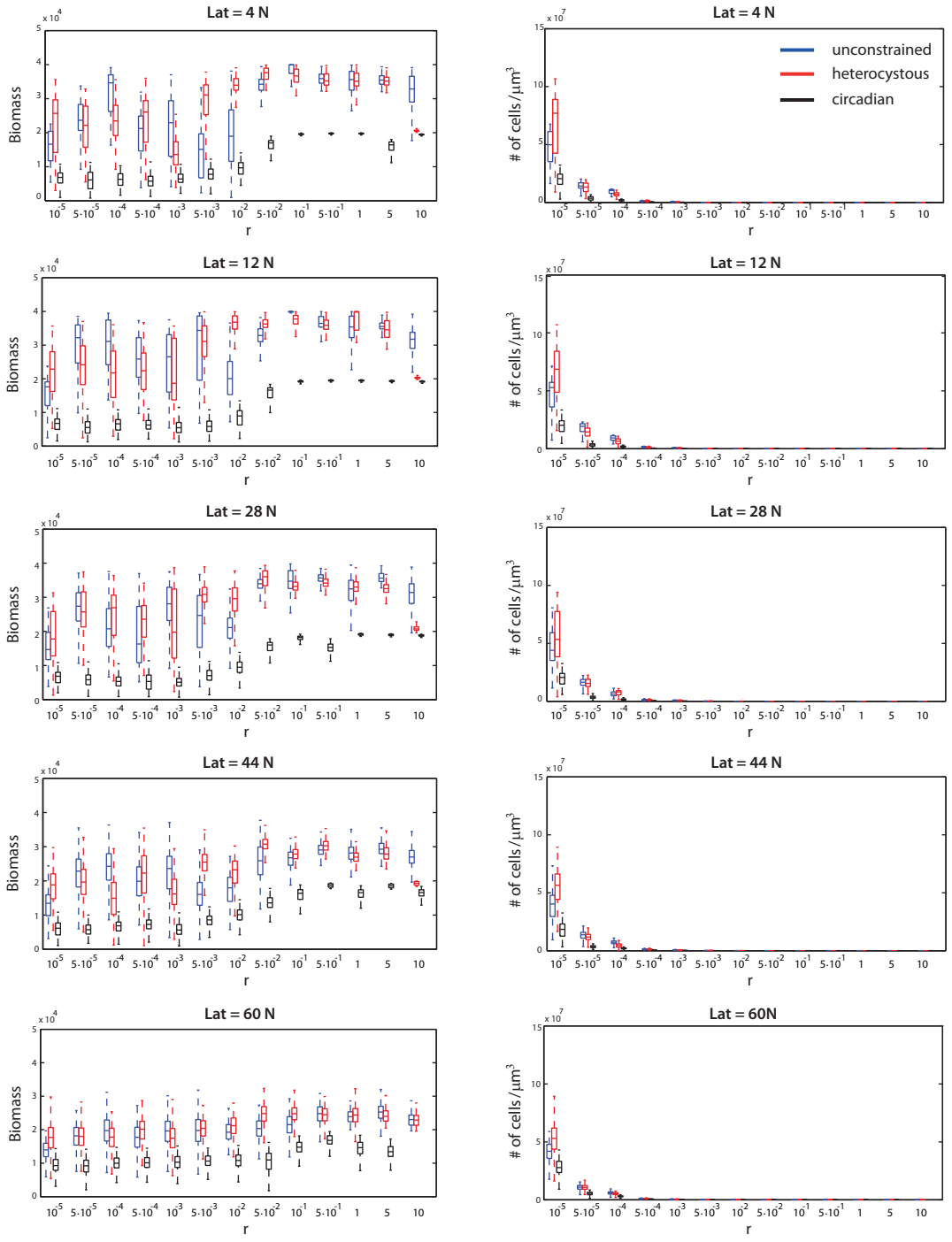


Figure 4.3: Box plot of the biomass production (left column) and population size (right column) computed daily over a simulation of a 35 year period at different latitudes and values of cell investment r. Biomass is expressed in fmol/μm^3. In each box, the central mark is the median, the horizontal edges of the box are the 25th and 75th percentiles and the whiskers extend to the most extreme data points not considered outliers (outliers not shown).

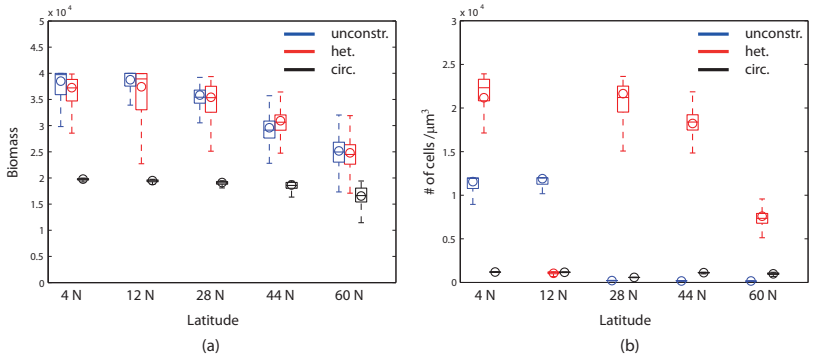


Figure 4.4: (a) Comparison of the box plots corresponding to the (a, r) providing the highest biomass at each latitude for each species. Parameter r represents the resource investment into reproduction, a is the resource investment in N-fixation. Both r and a are measured in $fmol \cdot cell^{-1} \cdot h^{-1}$ and biomass is expressed in fmol/μm^3. Data correspond to daily biomass over a 35 year period. (b). Box plot of the number of cells over a 35 year period obtained with the same (a, r) values as in panel (a). The open circle in each box plot corresponds to the mean values of biomass and number of cells, respectively.

with the optimal pairs at each latitude. Open circles indicate the mean biomass values used to determine the corresponding optimal pair. Figure 4.4b shows the population size of the two species obtained at each latitude with the same optimal (a, r) pairs determined for the biomass. At optimal (a, r) values, heterocystous and unconstrained species reach comparable levels of biomass production at every latitude. The biomass of circadian species is at lower levels in all cases. For the same (a,r) pairs however, the heterocystous species reach the highest number of cells except for latitude 12N. At latitudes 4N and 12N, the number of cells of the unconstrained species exceeds that of circadian species.

4.3.4 Best cell investment for biomass and population size.

Figure 4.5 shows a grid of the 5 values of a and the 13 values of r used in the simulations. The presence of a filled dot on a grid point indicates that a species (type indicated by the color code) at a given latitude reaches its best performance in terms of biomass at the corresponding (a, r) pair. The (a, r) pair values and the corresponding latitudes are listed in Table 4.2. The biomass of circadian and

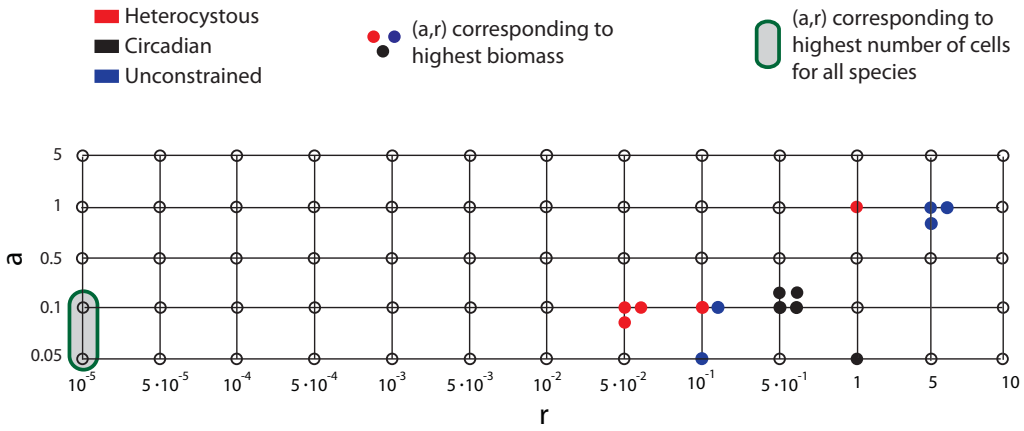


Figure 4.5: Best cell investment for each species and latitude. Colored dots indicate that the corresponding grid point is an optimal (a, r) pair that provides the highest biomass to a given species (see color code) at a given latitude (see Table 4.2).

heterocystous species is optimal for low values of a and high values of r. For the unconstrained species, the range of optimal a and r values is slightly bigger than that of the other two species. On the whole, each species has the highest biomass production for at most three grid points. In terms of population size, the filled oval on the left indicates that at any given latitude all three species reach their highest number of cells for one of the two included grid points.

4.3.5 Tradeoff between biomass and population size.

From Figure 4.3 it can be seen that for all species populations, biomass production increases with increasing r, while the number of cells decreases with increasing r. In Figure 4.4, species that have comparable biomass production for a given optimal (a, r) pair can have very different population sizes for the same a and r values. Figure 4.5 summarizes these findings showing that the (a, r) pairs corresponding to the highest biomass always differ from those corresponding to the highest population size. We can hence infer the existence of a tradeoff between biomass production and population growth in terms of number of cells. In order to mathematically verify this hypothesis, we computed the steady state population size V^* for system (4.1) and any given r, denoted as $V^*(r)$. We allow r to vary and

	Unconstrained	Circadian	Heterocystous
Lat 4N	$(0.1\ ,\ 5\cdot10^{-1})$	$(0.05\ ,\ 10^{-1})$	$(0.1\ ,\ 5\cdot10^{-2})$
Lat 12N	$(0.1\ ,\ 5\cdot10^{-1})$	$(0.1\ ,\ 10^{-1})$	$(1\ ,\ 1)$
Lat 28N	$(0.05\ ,\ 1)$	(1,5)	$(0.1\ ,\ 5\cdot10^{-1})$
Lat 44N	$(0.1\ ,\ 5\cdot10^{-1})$	$(1\ ,\ 5)$	$(0.1\ ,\ 5\cdot10^{-1})$
Lat 60N	$(0.1\ ,\ 5\cdot10^{-1})$	$(1\ ,\ 5)$	$(0.1\ ,\ 10^{-1})$

Table 4.2: Best (a,r) pair that provides the highest average biomass at each latitude.

fix values of irradiance, the functions δ_c and δ_n, and all other parameters (listed in Appendix C). The same approach would be also valid for systems (4.3) and (4.4). We then computed the sensitivity coefficient $\sigma = \sigma_V/\sigma_r$, where $\sigma_V = \Delta_{V^*(r)}/V^*(r)$ and $\sigma_r = \Delta_r/r$. The value of σ is an indicator of response of population size to the variation of r. For the range of r values considered in this model, we found that $-1 < \sigma < 0$ and that σ decreases towards -1 with increasing r. When $\sigma \sim -1$, the absolute value of the proportional change in V^* equals that of the proportional change in r, hence no tradeoff is expected. When the value of σ is far from -1, the response of the population size V^* to changes in r is weaker, indicating the existence of a tradeoff. This implies that the decrease of biomass per cell due to a smaller r cannot be compensated linearly by an increase of the population size. Details are provided in Appendix C.

4.4 Discussion

Long term population dynamics. The sample runs in Figure 4.2 show that the long term population dynamics superficially resemble a stochastic system, although no stochastic term is present in our models. Each species continuously tries to catch up with the external light availability, and has a different population size at the beginning of each new seasonal cycle. The irregular oscillations and the isolated peaks of the population size are suggesting chaotic dynamics and could be part of the explanation for the difficult predictability of cyanobacterial blooms observed in many bodies of water.

4.4.1 Cell differentiation and nitrogenase sensitivity to oxygen.

The unconstrained model represents a hypothetical, ideal combination of nitrogen and carbon fixation. In this hypothetical species, the nitrogenase enzyme is not sensitive to oxygen, allowing for an unconstrained coexistence of the two processes. Conversely, the circadian model simulated in this work is subject to very strict constraints. The inhibitory functions δ_c and δ_n stop carbon and nitrogen fixation during night and day respectively. Heterocystous species solve this biochemical incompatibility problem by developing oxygen protected cells in which nitrogenase can be constantly expressed. The results shown in Figures 4.3 and 4.4 indicate that heterocystous and unconstrained species have comparable performance in terms of biomass production. For every value of r in Figure 4.3a and for every optimal (a, r) pair in Figure 4.4a , the biomass value range of the two species overlap. The biomass of circadian species is instead considerably lower, although the value ranges partially overlap with that of the other species at latitude 60 N and for small values of r at lower latitudes (Figure 4.3a). These outcomes indicate that cell differentiation is a successful way to circumvent the biochemical constraints, since it allows for a biomass production that is comparable with the ideal unconstrained species. Hence, in the absence of exogenous sources of fixed nitrogen, spatial division of labor can hence significantly improve biomass accumulation. The scarce performance of circadian species can be explained by the fact that in this model, the endogenous rhythm and associated inhibition cycle is very strict and does not allow for coexistence between carbon and nitrogen fixation.

4.4.2 Tradeoff between biomass production and population growth.

As shown in the Results section, there is a tradeoff between biomass production and population size whereby given a value of carbon and nitrogen consumption, cyanobacteria can have either a high biomass output or a large population size, but not both at the same time. Interestingly, the values of r at which the tradeoff is higher (i.e. $r \leq 1$ and $-1 << \sigma < 0$, see Figure C.3) correspond to the domain

in which increasing r will lead to higher biomass production. This explains why in Figure 4.5, the optimal biomass production of heterocystous and circadian species is not at the largest r value ($r = 10$), but rather in the range $5 \cdot 10^{-2} \leq r \leq 1$. Above a certain value of r, the increase of the cell investment does not lead to an increase in biomass anymore. The reasons for the presence of the tradeoff can be found in the form of the function $Z = Z(N, C)$ regulating resource uptake and cell division. Since Z is a saturation function of carbon and nitrogen, excess nutrients can be incorporated into cell division only up to a certain level, above which population growth (cell division) cannot be increased by decreasing the cell investment.

4.4.3 Effect of latitude.

In all models, the main effect of latitude is observed in terms of biomass. Figures 4.3 and 4.4a show that biomass decreases at higher latitudes such as 44N and 60N. Moreover, at the higher latitudes, the difference between the biomass value ranges of the three species is also reduced. Hence at northern latitudes, the advantages provided by cellular differentiation over a strict circadian rhythm are less evident.

4.4.4 Conclusions.

This work shows that in cyanobacteria, spatial division of labor can overcome the biochemical constraints due to nitrogenase oxygen sensitivity and enhance biomass production. In terms of success at biomass accumulation, heterocystous species can become comparable to an idealized species that has no biochemical constraints. A strict temporal division of labor on the other hand leads to a lower biomass productivity. This suggests that in circadian species that lack outside sources of fixed nitrogen, selection would have to act to reduce the sensitivity of nitrogenase enzyme to oxygen in order to increase biomass productivity. This would allow for a more flexible temporal regulation of carbon and nitrogen fixation. However, we also show that due to a tradeoff between population size and biomass, different species can be successful in terms of different performance measures. This argument could

potentially be involved in explanations for the coexistence of different species types in the same environment. Finally, the oscillating dynamics of the population size could open new perspectives in the study of cyanobacterial blooms.

5 Conclusions

The evolution of multicellularity and cellular differentiation is unlikely to be explained with a unique argument. However, the study of simple multicellular organisms can help to disentangle some of the evolutionary forces influencing multicellularity and cellular differentiation. To this aim, this thesis investigated the developmental strategies of cyanobacteria as model organism.

By considering filamentous multicellular bacteria, we found that the emergence of multicellular life cycles can be driven by ecological factors such as birth and death rates of the cells (Chpt. 2). According to these rates, the distribution of filament lengths in the population varies at different population growth stages. The differences in filament lengths however does not necessarily correspond to differences in fitness. This last finding can lead to a paradigm modification: the evolution of life cycles dominated by multicellularity is not only due to fitness advantages of morphological features.

In the evolutionary history of a species, the developmental steps that were not undertaken can be also relevant, as shown in Chpt. 3. In cyanobacteria, a step that is physiologically possible is missing. This step corresponds to division of labor in a unicellular context, represented by communities of differentiated single cells. By modeling and comparing this hypothetical scenario with the extant multicellular one, we could understand the evolutionary conditions leading to the differentiation programs observed in natural populations. The non-multicellular differentiated species could neither assure resistance against disruptive mutations nor an optimization of the division of labor.

The different developmental programs can have specific roles such as the resolution of chemical incompatibilities at the intracellular level. In cyanobacteria, temporal and spatial division of labor are employed to prevent the inhibition of the nitrogenase enzyme due to oxygen. In Chpt. 4 we showed that spatial division

of labor can enhance biomass production up to a level comparable to that of a species where the biochemical constraints are absent.

This thesis demonstrates that the study of bacterial populations and their developmental strategies can deepen our understanding of the primordial evolution of multicellularity and cellular differentiation during the history of life.

Bibliography

[1] M. Ackermann, B. Stecher, N. E. Freed, P. Songhet, W. D. Hardt, and M. Doebeli. Self-destructive cooperation mediated by phenotypic noise. *Nature*, 454(7207):987–990, Aug 21 2008.

[2] M. Ackermann, S. C. Stearns, and U. Jenal. Senescence in a bacterium with asymmetric division. *Science*, 300(5627):1920, Jun 20 2003.

[3] F. Adamec, D. Kaftan, and L. Nedbal. Stress-induced filament fragmentation of Calothrix elenkinii (Cyanobacteria) is facilitated by death of high-fluorescence cells. *Journal of Phycology*, 41(4):835–839, Aug 2005.

[4] D. G. Adams. Heterocyst formation in cyanobacteria. *Current Opinion in Microbiology*, 3(6):618–624, 2000.

[5] N. S. R. Agawin, S. Rabouille, M. J. W. Veldhuis, L. Servatius, S. Hol, H. M. J. van Overzee, and J. Huisman. Competition and facilitation between unicellular nitrogen-fixing cyanobacteria and non-nitrogen-fixing phytoplankton species. *Limnology and Oceanography*, 52(5):2233–2248, Sep 2007.

[6] B. Amard and J. Bertrand-Sarfati. Microfossils in 2000 Ma old cherty stromatolites of the Franceville Group, Gabon. *Precambrian Research*, 81(3-4): 197–221, Feb 1997.

[7] C. S. O. Attolini and P. F. Stadler. Evolving towards the hypercycle: A spatial model of molecular evolution. *Physica D-Nonlinear Phenomena*, 217 (2):134–141, May 15 2006.

[8] R. Axelrod and W. D. Hamilton. The evolution of cooperation. *Science*, 211 (4489):1390–1396, 1981.

[9] L. A. Bach, T. Helvik, and F. B. Christiansen. The evolution of n-player cooperation–threshold games and ESS bifurcations. *Journal of Theoretical Biology*, 238:426, 2006.

[10] B. Bergman, J. R. Gallon, A. N. Rai, and L. J. Stal. N-2 fixation by non-heterocystous cyanobacteria. *FEMS Microbiology Reviews*, 19(3):139–185, February 1997.

[11] J. E. Berleman and J. R. Kirby. Deciphering the hunting strategy of a bacterial wolfpack. *FEMS Microbiology Reviews*, 33(5):942–957, 2009.

[12] I. Berman-Frank, P. Lundgren, and P. Falkowski. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Research in Microbiology*, 154(3):157 – 164, 2003.

[13] I. Berman-Frank, P. Lundgren, Y. B. Chen, H. Kupper, Z. Kolber, , B. Bergman, and P. Falkowski. Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium Trichodesmium. *Science*, 294 (5546):1534–1537, Nov 2001.

[14] I. Berman-Frank, K. D. Bidle, L. Haramaty, and P. G. Falkowski. The demise of the marine cyanobacterium, Trichodesmium spp., via an autocatalyzed cell death pathway. *Limnology and Oceanography*, 49(4):997–1005, Jul 2004.

[15] M. C. Boerlijst and P. Hogeweg. Spiral wave structure in pre-biotic evolution - hypercycles stable against parasites. *Physica D*, 48(1):17–28, Feb 1991.

[16] J. T. Bonner. *On development: the biology of form.* Harvard University Press, 1974.

[17] J. T. Bonner. The origins of multicellularity. *Integrative Biology*, 1(1):27, 1998.

[18] J. T. Bonner. *First signals: the evolution of multicellular development.* Princeton Univ. Press, Princeton, 2000.

[19] J. J. Boomsma and N. R. Franks. Social insects: from selfish genes to self organisation and beyond. *Trends in Ecology and Evolution*, 21(6):303 – 308, 2006.

[20] M. A. Brockhurst, A. Buckling, D. Racey, and A. Gardner. Resource supply and the evolution of public-goods cooperation in bacteria. *BMC Biology*, 6, May 14 2008.

[21] J. J. Brocks, R. Buick, R. E. Summons, and G. A. Logan. A reconstruction of archean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old mount bruce supergroup, hamersley basin, western australia. *Geochimica et Cosmochimica Acta*, 67(22):4321 – 4335, 2003.

[22] J. J. Brocks, G. A. Logan, R. Buick, and R. E. Summons. Archean molecular fossils and the early rise of eukaryotes. *Science*, 285(5430):1033–1036, Aug 13 1999.

[23] R. Buick. The antiquity of oxygenic photosynthesis: evidence from stromatolites in sulphate-deficient archaean lakes. *Science*, 255(5040):74–77, 1992.

[24] L. W. Buss. Somatic-cell parasitism and the evolution of somatic tissue compatibility. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, 79(17):5337–5341, 1982.

[25] L. W. Buss. Evolution, development, and the units of selection. *Proceedings of the National Academy of Sciences of the United States of America*, 80(5), 1983.

[26] L. W. Buss. *The evolution of individuality*. Princeton University Press, Princeton, NJ, 1987.

[27] D. G. Capone, J. A. Burns, J. P. Montoya, A. Subramaniam, C. Mahaffey, T. Gunderson, AF A. F. Michaels, and E. J. Carpenter. Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global BiogeochemicaL Cycles*, 19 (2), Jun 8 2005.

[28] W. W. Carmichael. Cyanobacteria secondary metabolites - The cyanotoxins. *Journal of Applied Bacteriology*, 72(6):445–459, Jun 1992.

[29] W. W. Carmichael. The cyanotoxins. volume 27 of *Advances in Botanical Research*, pages 211–256. 1997.

[30] E. J. Carpenter and C. C. Price. Marine oscillatoria (*Trichodesmium*): explanation for aerobic nitrogen fixation without heterocysts. *Science*, 191 (4233):1278–1280, 1976.

[31] E. J. Carpenter, J. Chang, M. Cottrell, J. Schubauer, H. W. Paerl, B. M. Bebout, and D. G. Capone. Reevaluation of nitrogenase oxygen-protective mechanisms in the planktonic marine cyanobacterium *Trichodesmium*. *Marine Ecology-Progress Series*, 65(2):151–158, Aug 1990.

[32] S. B. Carroll. Chance and necessity: the evolution of morphological complexity and diversity. *Nature*, 409:1102–1109, 2001.

[33] B. Charlesworth. *Evolution in age-structured populations*. Cambridge University Press, second edition, 1994.

[34] O. Choi, A. Das, C. P. Yu, and Z. Hu. Nitrifying bacterial growth inhibition in the presence of algae and cyanobacteria. *Biotechnology and Bioengineering*, 107(6):1004–1011, Dec 15 2010.

[35] M. J. Church, K. M. Bjorkman, D. M. Karl, M. A. Saito, and J. P. Zehr. Regional distributions of nitrogen-fixing bacteria in the Pacific Ocean. *Limnology and Oceanography*, 53(1):63–77, Jan 2008.

[36] M. J. Daft and W. D. P. Stewart. Light and electron-microscope observations on algal lysis by bacterium CP-1. *New Phytologist*, 72(4):799–808, 1973.

[37] A. de los Rios, C. Ascaso, J. Wierzchos, E. Fernandez-Valiente, and A. Quesada. Microstructural characterization of cyanobacterial mats from the McMurdo Ice Shelf, Antarctica. *Applied and Environmental Microbiology*, 70 (1):569–580, Jan 2004.

[38] M. Doebeli and N. Knowlton. The evolution of interspecific mutualisms. *Proceedings of the National Academy of Sciences*, 95(15):8676–8680, 1998.

[39] M. Dworkin. The myxobacteria: New directions in studies of procaryotic development. *Critical Reviews in Microbiology*, 1(4):435–452, 1972.

[40] M. Eigen. Selforganization of matter and evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465, 1971.

[41] M. Eigen and P. Schuster. Hypercycle - principle of natural self-organization .a. emergence of hypercycle. *Naturwissenschaften*, 64(11):541–565, 1977.

[42] M. Eigen and P. Schuster. Hypercycle - principle of natural self-organization .b. abstract hypercycle. *Naturwissenschaften*, 65(1):7–41, 1978.

[43] M. Eigen, W. C. Gardiner, and P. Schuster. Hypercycles and compartments - compartments assists - but do not replace - hypercyclic organization of early genetic information. *Journal of Theoretical Biology*, 85(3):407–411, 1980.

[44] R. El-Shehawy, C. Lugomela, A. Ernst, and B. Bergman. Diurnal expression of hetR and diazocyte development in the filamentous non-heterocystous cyanobacterium *Trichodesmium erythraeum*. *Microbiology-SGM*, 149(Part 5):1139–1146, May 2003.

[45] M. Enquist and O. Leimar. The evolution of cooperation in mobile organisms. *Animal Behaviour*, 45(4):747–757, Apr 1993.

[46] A. M. Evans, J. R. Gallon, A. Jones, M. Staal, L. J. Stal, M. Villbrandt, and T. J. Walton. Nitrogen fixation by Baltic cyanobacteria is adapted to the prevailing photon flux density. *New Phytologist*, 147(2):285–297, Aug 2000.

[47] R. Ferriere and R. E. Michod. The evolution of cooperation in spatially heterogeneous populations. *American Naturalist*, 147(5):692–717, May 1996.

[48] R. Ferriere, J. L. Bronstein, S. Rinaldi, R. Law, and M. Gauduchon. Cheating and the evolutionary stability of mutualisms. *Proceedings of the Royal Society of London Series B*, 296:773–780, 2002.

[49] M. Filippini, M. Svercel, E. Laczko, A. Kaech, U. Ziegler, and H. C. Bagheri. *Fibrella aestuarina* gen. nov., sp. nov., a filamentous bacterium of the family cytophagaceae isolated from north sea tidal flats and emended description of the genus *Rudanella* weon et al., 2008. *International Journal of Systematic and Evolutionary Microbiology*, 61:184–189, 2011.

[50] M. Filippini, A. Kaech, U. Ziegler, and H. C. Bagheri. *Fibrisoma limi* gen. nov., sp. nov., an orange pigmented filamentous bacterium of the family cytophagaceae isolated from north sea tidal flats. *International Journal of Systematic and Evolutionary Microbiology*, In press.

[51] J. A. Finzi-Hart, J. Pett-Ridge, P. K. Weber, R. Popa, S. J. Fallon, T. Gunderson, I. D. Hutcheon, K. H. Nealson, and D. G. Capone. Fixation and fate of C and N in the cyanobacterium *Trichodesmium* using nanometer-scale secondary ion mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, 106(15):6345–6350, Apr 2009.

[52] E. Flores, A. Herrero, C. P. Wolk, and I. Maldener. Is the periplasm continuous in filamentous multicellular cyanobacteria? *Trends in Microbiology*, 14 (10):439–443, Oct 2006.

[53] M. J. Follows, S. Dutkiewicz, S. Grant, and S. W. Chisholm. Emergent biogeography of microbial communities in a model ocean. *Science*, 315(5820): 1843–1846, 2007.

[54] J. F. Fontanari, M. Santos, and E. zathmary. Coexistence and error propagation in pre-biotic vesicle models: A group selection approach. *Journal of Theoretical Biology*, 239(2):247 – 256, 2006.

[55] S. A. Frank. Kin selection and virulence in the evolution of protocells and parasites. *Proceedings of the Royal Society of London Series B*, 258(1352): 153–161, Nov 22 1994.

[56] C. Fredriksson and B. Bergman. Nitrogenase quantity varies diurnally in a subset of cells within colonies of the nonheterocystous cyanobacteria *Trichodesmium* Spp. *Microbiology-UK*, 141(Part 10):2471–2478, Oct 1995.

[57] J. R. Gallon. Reconciling the Incompatible - N-2 Fixation and O-2. *New Phytologist*, 122(4):571–609, Dec 1992.

[58] M. Gantar and Z. Svircev. Microalgae and cyanobacteria: Food for thought. *Journal of Phycology*, 44(2):260–268, Apr 2008.

[59] F. Garcia-Pichel, A. Lopez-Cortes, and U. Nubel. Phylogenetic and morphological diversity of cyanobacteria in soil desert crusts from the Colorado plateau. *Applied And Environmental Microbiology*, 67(4):1902–1910, 2001.

[60] S. Gavrilets. Rapid transition towards the division of labor via evolution of developmental plasticity. *PLoS Comput Biol*, 6, 06 2010.

[61] O. M. Gilbert, K. R. Foster, N. J. Mehdiabadi, J. E. Strassmann, and D. C. Queller. High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proceedings of The National Academy of Science of the United States of America*, 104(21):8913–8917, May 22 2007.

[62] S. J. Giovannoni, S. Turner, G. J. Olsen, S. Barns, D. J. Lane, and N. R. Pace. Evolutionary relationships among cyanobacteria and green chloroplasts. *Journal Of Bacteriology*, 170(8):3584–3592, 1988.

[63] A. S. Griffin, S. A. West, and A. Buckling. Cooperation and competition in pathogenic bacteria. *Nature*, 430(7003):1024–1027, Aug 26 2004.

[64] R. K. Grosberg and R. R. Strathmann. The evolution of multicellularity: A minor major transition? *Annual Review of Ecology, Evolution, and Systematics*, 38(1):621–654, 2007.

[65] H. Guede. Grazing by protozoa as selection factor for activated sludge bacteria. *Microbial Ecology*, 5(3):225–237, 1979.

[66] M. W. Hahn and M. G. Hoefle. Grazing pressure by a bacterivorous flagellate reverses the relative abundance of comamonas acidovorans PX54 and vibrio strain CB5 in chemostat cocultures. *Applied and Environmental Microbiology*, 64(5):1910–1918, May 1998.

[67] M. W. Hahn, E. R. Moore, and M. G. Hoefle. Bacterial filament formation, a defense mechanism against flagellate grazing, is growth rate controlled in

bacteria of different phyla. *Applied and Environmental Microbiology*, 65(1): 25–35, 1999.

[68] W. D. Hamilton. The genetical evolution of social behaviour. i. *Journal of Theoretical Biology*, 7(1):1–&, July 1964.

[69] W. D. Hamilton. The genetical evolution of social behaviour. ii. *Journal of Theoretical Biology*, 7(1):17–&, July 1964.

[70] P. Hammerstein. Why is reciprocity so rare in social animals? a protestant appeal. In P. Hammerstein, editor, *Genetic and Cultural Evolution of Cooperation*, Dahlem Workshop Reports: Environmental Sciences Research Report, pages 83–93, 2003. 90th Dahlem Workshop on Genetic and Cultural Evolution of Cooperation, Berlin, Germany, Jun 23-28, 2002.

[71] B. P. Han. Photosynthesis-irradiance response at physiological level: a mechanistic model. *Journal of Theoretical Biology*, 213(2):121, 2001.

[72] G. Hardin. The tragedy of the commons. *Science*, 162:143, 1968.

[73] M. Hochberg, D. Rankin, and M. Taborsky. The coevolution of cooperation and dispersal in social groups and its implications for the emergence of multicellularity. *BMC Evolutionary Biology*, 8(1):238, 2008.

[74] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

[75] J. Hofbauer, P. Schuster, and K. Sigmund. Evolutionary stable strategies and game dynamics. *Journal of Theoretical Biology*, 81(3):609–612, 1979.

[76] H. J. Hofmann. Precambrian Microflora, Belcher Islands, Canada - Significance and Systematics. *Journal of Paleontology*, 50(6):1040–1073, 1976.

[77] P. Hogeweg and N. Takeuchi. Multilevel selection in models of prebiotic evolution: Compartments and spatial self-organization. *Origins of Life and Evolution of Biospheres*, 33(4):375, 2003.

[78] D. Honda, A. Yokota, and J. Sugiyama. Detection of seven major evolutionary lineages in cyanobacteria based on the 16s rRNA gene sequence analysis with new sequences of five marine synechococcus strains. *Journal of Molecular Evolution*, 48(6):723–739, 1999.

[79] P. Hyenstrand, P. Blomqvist, and A. Petersson. Factors determining cyanobacterial success in aquatic systems - a literature review. In Forsberg, C and Pettersson, K, editor, *Advances in Limnology 51 - Lake Erken -*

50 Years of limnological Research , volume 51 of *Ergebnisse der Limnologie*, pages 41–62. 1998.

[80] H .C .P .Matthijs J .Huisman and P .M .Visser, editors. *Harmful Cyanobacteria*. Springer Netherlands, 2005.

[81] S. Janson, E. J. Carpenter, and B. Bergman. Compartmentalization of nitrogenase in a nonheterocystous cyanobacterium *Trichodesmium Contortum*. *FEMS Microbiology Letters*, 118(1-2):9–14, May 1 1994.

[82] P. Jordan, P. Fromme, H. T. Witt, O. Klukas, W. Saenger, and N. Krauss. Three-dimensional structure of cyanobacterial photosystem I at 2.5 angstrom resolution. *Nature*, 411(6840):909–917, Jun 21 2001.

[83] K. Jurgens, H. Arndt, and K. O. Rothhaupt. Zooplankton-mediated changes of bacterial community structure. *Microbial Ecology*, 27(1):27–42, Jan-feb 1994.

[84] H. Kageyama, T. Nishiwaki, M. Nakajima, H. Iwasaki, T. Oyama, and T. Kondo. Cyanobacterial circadian pacemaker: Kai protein complex dynamics in the kaic phosphorylation cycle in vitro. *Molecular Cell*, 23(2): 161–171, Jul 21 2006.

[85] M. Kahru, U. Horstmann, and O. Rud. Satellite detection of increased cyanobacteria blooms in the Baltic Sea - natural fluctuation or ecosystem change. *Ambio*, 23(8):469–472, Dec 1994.

[86] D. Kaiser. Building a multicellular organism. *Annual Review of Genetics*, 35(1):103–123, 2001.

[87] A. Kamp, H. Roy, and H. N. Schulz-Vogt. Video supported analysis of Beggiatoa filament growth, breakage, and movement. *Microbial Ecology*, 56 (3):484–491, Oct 2008.

[88] K. Kastovska, J. Elster, M. Stibal, and D. Santruckova. Microbial assemblages in soil microbial succession after glacial retreat in Svalbard (high Arctic). *Microbial Ecology*, 50(3):396–407, Oct 2005.

[89] R. A. Kay. Microalgae as Food and Supplement. *Critical Reviews in Food Science and Nutrition*, 30(6):555–573, 1991.

[90] C. N. Keim, J. L. Martins, F. Abreu, A. S. Rosado, H. L. de Barros, R. Borojevic, U. Lins, and M. Farina. Multicellular life cycle of magnetotactic prokaryotes. *FEMS Microbiology Letters*, 240(2):203–208, Nov 15 2004.

[91] T. Killingback, J. Bieri, and T. Flatt. Evolution in group-structured populations can resolve the tragedy of the commons. *Proceedings of the Royal Society B*, 273(1593):1477–1481, 2006.

[92] N. King. The unicellular ancestry of animal development. *Developmental Cell*, 7(3):313–325, 2004.

[93] D. L. Kirk. Seeking the ultimate and proximate causes of Volvox multicellularity and cellular differentiation. *Integrative and Comparative Biology*, 43 (2):247–253, Apr 2003.

[94] D. L. Kirk. A twelve-step program for evolving multicellularity and a division of labor. *Bioessays*, 27(3):299–310, Mar 2005.

[95] A. H. Knoll. The Multiple Origins of Complex Multicellularity. In R. Jeanloz and K. H. Freeman, editor, *Annual Review of Earth and Planetary Sciences, Vol 39*, volume 39 of *Annual Review of Earth and Planetary Sciences*, pages 217–239. 2011.

[96] T. Kondo, C. A. Strayer, R. D. Kulkarni, W. Taylor, M. Ishiura, S. S. Golden, and C. H. Johnson. Circadian rhythms in prokaryotes - luciferase as a reporter of circadian gene expression in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 90(12):5672–5676, Jun 15 1993.

[97] K. O. Konhauser. Deepening the early oxygen debate. *Nature Geoscience*, 2(4):241–242, Apr 2009.

[98] J. U. Kreft. Biofilms promote altruism. *Microbiology*, 150(8):2751–2760, 2004.

[99] J. Kromkamp. Formation and functional Significance of Storage Products in Cyanobacteria. *New Zealand Journal of Marine and Freshwater Research*, 21(3):457–465, 1987.

[100] M. Kruskopf. Growth and filament length of the bloom forming *Oscillatoria simplicissima* (oscillatoriales, cyanophyta) in varying N and P concentrations. *Hydrobiologia*, 556(1):357–362, 2006.

[101] T. Kuritz and C. P. Wolk. Use of filamentous cyanobacteria for biodegradation of organic pollutants. *Applied and Environmental Microbiology*, 61(1): 234–238, Jan 1995.

[102] G. Kurosawa, K. Aihara, and Y. Iwasa. A model for the circadian rhythm of cyanobacteria that maintains oscillation without gene expression. *Biophysical Journal*, 91(6):2015–2023, 2006.

[103] C. Lanave, G. Preparata, C. Saccone, and G. Serio. A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution*, 20(1):86–93, 1984.

[104] D. W. Lawlor. *Photosynthese*. Georg Thieme Verlag, Stuttgart, 1990.

[105] L. Lehmann and L. Keller. The evolution of cooperation and altruism - a general framework and a classification of models. *Journal of Evolutionary Biology*, 19(5):1365–1376, Sep 2006.

[106] J. Lehtimaki, P. Moisander, K. Sivonen, and K. Kononen. Growth, nitrogen fixation, and nodularin production by two baltic sea cyanobacteria. *Applied and Environmental Microbiology*, 63(5):1647–1656, 1997.

[107] O. Leimar and P. Hammerstein. Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London Series B*, 268(1468): 745–753, Apr 7 2001.

[108] O. Leimar and P. Hammerstein. Facing the facts. *Journal of Evolutionary Biology*, 19(5):1403–1405, Sep 2006.

[109] P. H. Leslie. On the use of matrices in certain population mathematics. *Biometrika*, 33(3):183–212, 1945.

[110] K. Lewis. Programmed death in bacteria. *Microbiol. Mol. Biol. Rev.*, 64(3): 503–514, 2000.

[111] S. J. Lin, . Henze, P. Lundgren, B. Bergman, and E. J. Carpenter. Whole-cell immunolocalization of nitrogenase in marine diazotrophic cyanobacteria, Trichodesmium spp. *Applied and Environmental Microbiology*, 64(8):3052–3058, Aug 1998.

[112] E. Litchman. Competition and coexistence of phytoplankton under fluctuating light: experiments with two cyanobacteria. *Aquatic Microbial Ecology*, 31(3):241–248, Apr 3 2003.

[113] B. Loll, J. Kern, W. Saenger, A. Zouni, and J. Biesiadka. Towards complete cofactor arrangement in the 3.0 angstrom resolution structure of photosystem II. *Nature*, 438(7070):1040–1044, Dec 15 2005.

[114] M M. Travisano and G. J. Velicer. Strategies of microbial cheater control. *Trends in Microbiology*, 12(2):72–78, FEB 2004.

[115] V. Mariscal, A. Herrero, and E. Flores. Continuous periplasm in a filamentous, heterocyst-forming cyanobacterium. *Molecular Microbiology*, 65 (4):1139–1145, Aug 2007.

[116] G. Markou and D. Georgakakis. Cultivation of filamentous cyanobacteria (blue-green algae) in agro-industrial wastes and wastewaters: A review. *Applied Energy*, 88(10, SI):3389–3401, Oct 2011.

[117] M. R. Marler, P. Gehrman, J. L. Martin, and S. Ancoli-Israel. The sigmoidally transformed cosine curve: A mathematical model for circadian rhythms with symmetric non-sinusoidal shapes. *Statistics in Medicine*, 25 (22):3893–3904, Nov 30 2006.

[118] J. C. Meeks and J. Elhai. Regulation of cellular differentiation in filamentous cyanobacteria in free-living and plant-associated symbiotic growth states. *Microbiology and Molecular Biology Reviews*, 66(1):94–121, Mar 2002.

[119] R. E. Michod. Population biology of the 1st replicators - on the origin of the genotype, phenotype and organism. *American Zoologist*, 23(1):5–14, 1983.

[120] R. E. Michod. *Darwinian dynamics: evolutionary transitions in fitness and individuality*. Princeton Univ. Press, 2000.

[121] R. E. Michod. Evolution of individuality during the transition from unicellular to multicellular life. *Proceedings of the National Academy of Sciences*, 104(Suppl 1):8613–8618, 2007.

[122] R. E. Michod and D. Roze. Cooperation and conflict in the evolution of multicellularity. *Heredity*, 86(Part 1):1–7, Jan 2001.

[123] R. E. Michod, Y. Viossat, C. A. Solari, M. Hurand, and A. M. Nedelcu. Life-history evolution and the origin of multicellularity. *Journal of Theoretical Biology*, 239:257–272, 2006.

[124] A. Mitsui, S. Kumazawa, A. Takahashi, H. Ikemoto, S. Cao, and T. Arai. Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically. *Nature*, 323(6090):720–722, Oct 23 1986.

[125] F. M. Monteiro, M. J. Follows, and S. Dutkiewicz. Distribution of diverse nitrogen fixers in the global ocean. *Global Biogeochemical Cycles*, 24, Sep 10 2010.

[126] C. W. Mullineaux, V. Mariscal, A. Nenninger, H. Khanum, A. Herrero, E. Flores, and D. G. Adams. Mechanism of intercellular molecular exchange in heterocyst-forming cyanobacteria. *EMBO Journal*, 27(9):1299–1308, May 7 2008.

[127] M. Nakamaru, H. Matsuda, and Y. Iwasa. The evolution of cooperation in a lattice-structured population. *Journal of Theoretical Biology*, 184(1):65 – 81, 1997.

[128] P. H. Nielsen, C. Kragelund, R. J. Seviour, and J. L. Nielsen. Identity and ecophysiology of filamentous bacteria in activated sludge. *FEMS Microbiology Reviews*, 33(6):969–998, 2009.

[129] K. Niklas. The evolution of plant body plans — a biomechanical perspective. *Annals of Botany*, 85(4):411–438, 2000.

[130] S. B. Ning, H. L. Guo, L Wang, and Y. C. Song. Salt stress induces programmed cell death in prokaryotic organism Anabaena. *Journal of Applied Microbiology*, 93(1):15–28, 2002.

[131] W. T. De Nobel, H. C. P. Matthijs, E. Von Elert, and L. R. Mur. Comparison of the light-limited growth of the nitrogen-fixing cyanobacteria Anabaena and Aphanizomenon. *New Phytologist*, 138(4):579–587, Apr 1998.

[132] M. A. Nowak and R. M. May. Evolutionary games and spatial chaos. *Nature*, 359(6398):826–829, Oct 29 1992.

[133] M. A. Nowak, S. Bonhoeffer, and R. M. May. Spatial games and the maintenance of cooperation. *Proceedings of the National Academy of Sciences of the United States of America*, 91(11):4877–4881, May 24 1994.

[134] H. Ohtsuki, C. Hauert, E. Lieberman, and M. A. Nowak. A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441 (7092):502–505, May 25 2006.

[135] H. W. Paerl and B. M. Bebout. Direct Measurement of O-2-depleted Microzones in marine Oscillatoria - Relation to N-2 Fixation. *Science*, 241(4864): ul442–445, 1988.

[136] H. W. Paerl, J. L. Pinckney, and T. F. Steppe. Cyanobacterial-bacterial mat consortia: examining the functional unit of microbial survival and growth in extreme environments. *Environmental Microbiology*, 2(1):11–26, Feb 2000.

[137] F. Partensky, W. R. Hess, and D. Vaulot. Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiology And Molecular Biology Reviews*, 63(1):106–127, 1999.

[138] J. Pernthaler. Predation on prokaryotes in the water column and its ecological implications. *Nature Reviews Microbiology*, 3(7):537–546, 2005.

[139] J. Pernthaler, T. Posch, K. Simek, J. Vrba, R. Amann, and R. Psenner. Contrasting bacterial strategies to coexist with a flagellate predator in an experimental microbial assemblage. *Applied And Environmental Microbiology*, 63(2):596–601, 1997.

[140] J. Pernthaler, E. Zollner, F. Warnecke, and K. Jurgens. Bloom of filamentous bacteria in a mesotrophic lake: Identity and potential controlling mechanism. *Applied and Environmental Microbiology*, 70(10):6272–6281, 2004.

[141] T. Pfeiffer and S. Bonhoeffer. An evolutionary scenario for the transition to undifferentiated multicellularity. *Proceedings of the National Academy of Sciences*, 100(3):1095–1098, 2003.

[142] T. Pfeiffer, S. Schuster, and S. Bonhoeffer. Cooperation and competition in the evolution of ATP-producing pathways. *Science*, 292(5516):504–507, 2001.

[143] D. Posada and K. A. Crandall. MODELTEST: testing the model of dna substitution. *Bioinformatics*, 14(9):817–818, 1998.

[144] A. F. Post, J. G. Loogman, and L. R. Mur. Photosynthesis, carbon flows and growth of *Oscillatoria-Agardhii* Gomont in environments with a periodic supply of light. *Journal of General Microbiology*, 132(Part 8):2129–2136, Aug 1986.

[145] P. Prusinkiewicz and A. Lindenmayer. *The algorithmic beauty of plants*. Springer-Verlag, New York, 1990. ISBN 9780387972978.

[146] J. P. Zehr R. A. Foster, A. Subramaniam. Distribution and activity of diazotrophs in the eastern equatorial atlantic. *Environmental Microbiology*, 11(4):741, 2009.

[147] S. Rabouille, M. Staal, L. J. Stal, and K. Soetaert. Modeling the dynamic regulation of nitrogen fixation in the cyanobacterium *Trichodesmium* sp. *Applied and Environmental Microbiology*, 72(5):3217–3227, May 2006.

[148] P. B. Rainey and K. Rainey. Evolution of cooperation and conflict in experimental bacterial populations. *Nature*, 425(6953):72–74, September 2003.

[149] T. R. Ramothokang, G. D. Drysdale, and F. Bux. Isolation and cultivation of filamentous bacteria implicated in activated sludge bulking. *Water Sa*, 29 (4):405–410, Oct 2003.

[150] J. Raven and J. Allen. Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biology*, 4(3):209, 2003.

[151] K. C. Rice and K. W. Bayles. Molecular control of bacterial death and lysis. *Microbiology and Molecular Biology Reviews*, 72(1):85+, MAR 2008.

[152] R. Rippka, J. DeReuelles, J. B. Waterbury, M. Herdman, and R. Y. Stanier. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.*, 111:1–61, 1979.

[153] G. Roberts and T. N. Sherrat. Development of cooperative relationships through increasing investment. *Nature*, 394:175–179, 1998.

[154] F. Rodriguez, J. L. Oliver, A. Marin, and J. R. Medina. The general stochastic-model of nucleotide substitution. *Journal Of Theoretical Biology*, 142(4):485–501, 1990.

[155] G. Roeselerss, T. B. Norris, R. . Castenholz, S. Rysgaardn, R. N. Glud, M. Kuhl, and G. Muyzer. Diversity of phototrophic bacteria in microbial mats from Arctic hot springs (Greenland). *Environmental Microbiology*, 9 (1):26–38, Jan 2007.

[156] A. Rokas. The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annual Review of Genetics*, 42(1):235–251, 2008.

[157] F. Ronquist and J. P. Huelsenbeck. Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12):1572–1574, 2003.

[158] M. R. Rose. *Evolutionary Biology of Aging*. Oxford, 1991.

[159] D. A. Rosen, T. M. Hooton, W. E. Stamm, P. A. Humphrey, and S. J. Hultgren. Detection of intracellular bacterial communities in human urinary tract infection. *PLoS Medicine*, 4(12):1949–1958, Dec 2007.

[160] E. Rosenberg, K. H. Keller, and M. Dworkin. Cell density-dependent growth of myxococcus xanthus on casein. *Journal of Bacteriology*, 129(2):770–777, 1977.

[161] V. Rossetti, B. E. Schirrmeister, M. V. Bernasconi, and H. C. Bagheri. The evolutionary path to terminal differentiation and division of labor in cyanobacteria. *Journal of Theoretical Biology*, 262(1):23 – 34, 2010.

[162] M. R. Roussel, D. Gonze, and A. Goldbeter. Modeling the differential fitness of cyanobacterial strains whose circadian oscillators have different free-running periods: Comparing the mutual inhibition and substrate depletion hypotheses. *Journal of Theoretical Biology*, 205(2):321–340, Jul 21 2000.

[163] F. C. Rubio, F. G. Camacho, J. M. F. Sevilla, Y. Chisti, and E. M. Grima. A mechanistic model of photosynthesis in microalgae. *Biotechnology and Bioengeneering*, 81(4):459–473, 2003.

[164] J. L. Sachs, U. G. Mueller, T. P. Wilcox, and J. J. Bull. The evolution of cooperation. *The Quarterly Review of Biology*, 79(2):135–160, 2004.

[165] M. Saier and G. R. Jacobson. *The molecular basis of sex and differentiation.* Springer-Verlag, New York, 1984.

[166] G. Sandh, R. El-Shehawy, B. Diez, and B. Bergman. Temporal separation of cell division and diazotrophy in the marine diazotrophic cyanobacterium Trichodesmium erythraeum IMS101. *FEMS Microbiology Letters*, 295(2): 281–288, Jun 2009.

[167] J. Sardanyés and R. V. Solé. Bifurcations and phase transitions in spatially extended two-member hypercycles. *Journal of Theoretical Biology*, 243(4): 468 – 482, 2006.

[168] B. E. Schirrmeister, A. Antonelli, and H. C. Bagheri. The origin of multicellularity in cyanobacteria. *BMC Evolutionary Biology*, 11(45), 2011.

[169] W. Schleip. August Weismanns Bedeutung für die Entwicklung der Zoologie und allgemeinen Biologie zu seinem hundertsten Geburtstag am 17. januar 1934. *Naturwissenschaften*, 22(3), 1934.

[170] M. A. Schneegurt, D. M. Sherman, S. Nayar, and L. A. Sherman. Oscillating behavior of carbohydrate granule formation and dinitrogen fixation in the cyanobacterium *Cyanothece* sp. strain ATCC-51142. *Journal of Bacteriology*, 176(6):1586–1597, Mar 1994.

[171] J. W. Schopf. Microfossils of the early archean apex chert: New evidence of the antiquity of life. *Science*, 260(5108):640–646, 1993.

[172] J. W. Schopf. Disparate rates, differing fates: tempo and mode of evolution changed from the precambrian to the phanerozoic. *Proceedings of the National Academy of Sciences*, 91(15):6735–6742, 1994.

[173] J. W. Schopf. Fossil evidence of archaean life. *Philosophical Transactions of the Royal Society B*, 361(1470):869–885, 2006.

[174] J. W. Schopf and M. R. Walter. *Origin and early evolution in cyanobacteria: The geological evidence.* In: Carr NG and Whitton BA (eds) The biology of cyanobacteria, pp 543-564, Blackwell, Oxford, and University of California Press, Berkeley., 1982.

[175] J. W. Schopf, A. B. Kudrayavtsev, A. D. Czaja, and A. B. Tripathi. Evidence of archaean life: Stromatolites and microfossils. *Precambrian Research*, 158: 141–155, 2007.

[176] K. G. Sellner. Physiology, ecology, and toxic properties of marine cyanobacteria blooms. *Limnology and Oceanography*, 42(5, Part 2):1089–1104, Jul 1997.

[177] P. S. Seo and A. Yokota. The phylogenetic relationships of cyanobacteria inferred from 16srRNA, gyrB, rpoC1 and rpoD1 gene sequences. *The Journal of General and Applied Microbiology*, 49, 2003.

[178] J. A. Shapiro. Thinking about bacterial populations as multicellular organisms. *Annual Review of Microbiology*, 52(1):81–104, 1998.

[179] S. Shikano, L. S. Luckinbill, and Y. Kurihara. Changes of traits in a bacterial population associated with protozoal predation. *Microbial Ecology*, 20(1):75, 1990.

[180] D. C. Sigee, R Glenn, M. J. Andrews, E. G. Bellinger, R. D. Butler, H. A. S. Epton, and R. D. Hendry. Biological control of cyanobacteria: principles and possibilities. *Hydrobiologia*, 395:161–172, FEB 1999. Conference on Hydrobiologia, Leicester, England, Mar, 1996.

[181] J. M. Smith. *Evolution and the Theory of Games.* Cambridge University Press, 1982.

[182] J. M. Smith and G. R. Price. Logic of animal conflict. *Nature*, 246(5427): 15–18, 1973.

[183] J. M. Smith and E. Szathmáry. *The major transitions in evolution.* Freeman, Oxford, 1995.

[184] J. Maynard Smith. Group selection and kin selection. *Nature*, 201:11451147, 1964.

[185] J. Maynard Smith. Hypercycles and the origin of life. *Nature*, 280(5722): 445–446, 1979.

[186] C. A. Solari, S. Ganguly, J. O. Kessler, R. E. Michod, and R. E. Goldstein. Multicellularity and the functional interdependence of motility and molecular transport. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5):1353–1358, 2006.

[187] C. A. Solari, J. O. Kessler, and R. E. Michod. A hydrodynamics approach to the evolution of multicellularity: Flagellar motility and Germ-Soma differentiation in volvocalean green algae. *The American Naturalist*, 167(4): 537–554, 2006.

[188] M. Staal, F. J. R. Meysman, and L. J. Stal. Temperature excludes N-2-fixing heterocystous cyanobacteria in the tropical oceans. *Nature*, 425(6957):504–507, Oct 2 2003.

[189] L. J. Stal and W. E. Krumbein. Temporal separation of nitrogen-fixation and photosynthesis in the filamentous, nonheterocystous cyanobacterium oscillatoria sp. *Archives of Microbiology*, 149(1):76–80, NOV 1987.

[190] L. J. Stal and A. E. Walsby. The daily integral of nitrogen fixation by planktonic cyanobacteria in the Baltic Sea. *New Phytologist*, 139(4):665–671, Aug 1998.

[191] L. J. Stal, P. Albertano, B. Bergman, K. von Brockel, J. R. Gallon, P. K. Hayes, K. Sivonen, and A. E. Walsby. BASIC: Baltic Sea cyanobacteria. An investigation of the structure and dynamics of water blooms of cyanobacteria in the Baltic Sea - responses to a changing environment. *Continental Shelf Research*, 23(17-19):1695–1714, Nov-Dec 2003.

[192] R. Y. Stanier and G. Cohenbazire. Phototropic prokaryotes - cyanobacteria. *Annual Review of Microbiology*, 31:225–274, 1977.

[193] S. M. Stanley. An ecological theory for the sudden origin of multicellular life in the late precambrian. *Proceedings of the National Academy of Sciences of the United States of America*, 70(5):1486–1489, May 1973.

[194] J. E. Strassmann, Y. Zhu, and D. C. Queller. Altruism and social cheating in the social amoeba *Dictyostelium discoideum*. *Nature*, 408(6815):965–967, DEC 21 2000.

[195] S. Sultan. Phenotypic plasticity for plant development, function and life history. *Trends in Plant Science*, 5(12):537–542, 2000.

[196] S. Suzuki and E. Akiyama. Chaos, oscillation and the evolution of indirect reciprocity in n-person games. *Journal of Theoretical Biology*, 252:686, 2008.

[197] D. L. Swofford. PAUP*. phylogenetic analysis using parsimony (and other methods) version 4.0. *Sinauer Associates, Sunderland, Massachusetts.*, 2002.

[198] E. Szathmáry and L. Demeter. Group selection of early replicators and the origin of life. *Journal of Theoretical Biology*, 128(4):463–486, Oct 21 1987.

[199] G. W. Tannock. Analysis of the intestinal microflora: a renaissance. *Antonie van Leeuwenhoek International Journal of General and Molecular Microbiology*, 76(1-4):265–278, Nov 1999.

[200] J. D. Thompson, T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 25(24):4876–4882, 1997.

[201] D. Tilman. Resource competition between planktonic algae - experimental and theoretical approach. *Ecology*, 58(2):338–348, 1977.

[202] A. Tomitani, A. H. Knoll, C. M. Cavanaugh, and T. Ohno. The evolutionary diversification of cyanobacteria: Molecular-phylogenetic and paleontological perspectives. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 103(14):5442–5447, 2006.

[203] R. L. Trivers. Evolution of reciprocal altruism. *Quarterly Review of Biology*, 46(1):35–&, 1971.

[204] J. W. Tukey. Comparing individual means in the analysis of variance. *Biometrics*, 5(2):99–114, 1949.

[205] S. Turner, K. M. Pryer, V. P. W. Miao, and J. D. Palmer. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small submit rRNA sequence analysis. *Journal of Eukaryotic Microbiology*, 46(4): 327–338, 1999.

[206] E. J. van Hannen, G. Zwart, M. P. van Agterveld, H. J. Gons, J. Ebert, and H. J. Laanbroek. Changes in bacterial and eukaryotic community structure after mass lysis of filamentous cyanobacteria associated with viruses. *Applied and Environmental Microbiology*, 65(2):795–801, 1999.

[207] G. J. Velicer, L Kroos, and R. E. Lenski. Developmental cheating in the social bacterium *Myxococcus xanthus*. *Nature*, 404(6778):598+, Apr 6 2000.

[208] T. A. Vishnivetskaya. *Permafrost Soils*, chapter VI. Springer-Verlag Berlin, 2009.

[209] K. Vrede, M. Heldal, S. Norland, and G. Bratbak. Elemental composition (C, N, P) and cell volume of exponentially growing and nutrient-limited bacterioplankton. *Applied and Environmental Microbiology*, 68(6):2965–2971, 2002.

[210] A. E. Walsby. Cyanobacterial heterocysts: terminal pores proposed as sites of gas exchange. *Trends in Microbiology*, 15(8):340–349, aug 2007.

[211] M. G. Weinbauer. Ecology of prokaryotic viruses. *FEMS Microbiology Reviews*, 28(2):127–181, May 2004.

[212] A. Weismann. *Essays Upon Heredity*, chapter II,IV. Oxford At the Clarendon Press, 1889. http://www.esp.org:80/books/weismann/essays/facsimile/.

[213] H. Y. Weon, H. J. Noh, J. A. Son, H. B. Jang, B. Y. Kim, S. W. Kwon, and E. Stackebrandt. *Rudanella lutea* gen. nov., sp. nov., isolated from an air sample in Korea. *International Journal of Systematic and Evolutionary Microbiology*, 58(2):474–478, 2008.

[214] S. A. West, A. S. Griffin, A. Gardner, and S. P. Diggle. Social evolution theory for microorganisms. *Nature Reviews Microbiology*, 4(8):597–607, Aug 2006.

[215] F. Westall, C. E. J. de Ronde, G. Southam, N. Grassineau, M. Colas, C. S. Cockell, and H. Lammer. Implications of a 3.472-3.333 Gyr-old subaerial microbial mat from the Barberton greenstone belt, South Africa for the UV environmental conditions on the early Earth. *Philosophical Transactions of the Royal Society B*, 361(1474):1857–1875, Oct 29 2006.

[216] B. A. Whitton and M. Potts Editors. *The ecology of cyanobacteria. Their diversity in time and space*. Springer, 2000.

[217] D. E. Whitworth, editor. *Myxobacteria: Multicellularity and Differentiation*. ASM Press, Washington, 2008.

[218] M. Willensdorfer. On the evolution of differentiated multicellularity. *Evolution*, 63(2):306–323, January 2009.

[219] G. C. Williams. Pleiotropy, natural-selection, and the evolution of senescence. *Evolution*, 11(4):398–411, 1957.

[220] C. P. Wolk. Heterocyst formation. *Annual Review of Genetics*, 30:59–78, 1996.

[221] H. Wu, K. Gao, V. E. Villafane, T. Watanabe, and E. W. Helbling. Effects of solar UV radiation on morphology and photosynthesis of filamentous cyanobacterium Arthrospira platensis. *Applied and Environmental Microbiology*, 71(9):5004–5013, 2005.

[222] Q. L. Wu, J. Boenigk, and M. W. Hahn. Successful predation of filamentous bacteria by a nanoflagellate challenges current models of flagellate bacterivory. *Applied and Environmental Microbiology*, 70(1):332–339, 2004.

[223] J. P. Zehr, J. B. Waterbury, P. J. Turner, J. P. Montoya, E Omoregie, G. F. Steward, A Hansen, and D. M. Karl. Unicellular cyanobacteria fix N-2 in the subtropical North Pacific Ocean. *Nature*, 412(6847):635–638, Aug 2001.

A Supplementary material of Chpt. 2

A.1 Algorithm scheme

Figure A.1 illustrates schematically the algorithm used to simulate the dynamics of bacterial populations. The five steps are performed every iteration. In the first step, the population size is computed as the total number of cells, N_c. This quantity is used to compute the birth and death rates β and δ, that are density dependent. Secondly, a loop over all the filaments in the population begins. For each filament, the number of cell divisions and cell lyses are calculated and subsequently applied to the filament, that can elongate (step 3) and break (step 4). The resulting new individuals are stored as filaments of the next iteration (step 5).

A.2 Nonlinear birth and death rates

We tested the model with two different types of nonlinear birth and death rate functions. We used sigmoidal-shaped curves given by

$$\begin{aligned} \beta &= \frac{k_b}{k_b + N_c}, \quad k_b = \frac{\theta\, N_c^*}{1-\theta} \\ \delta &= \frac{N_c}{N_c + k_d}, \quad k_d = \frac{(1-\theta)\, N_c^*}{\theta} \end{aligned} \tag{A.1}$$

and hyperbolic functions of the form:

$$\begin{aligned} \beta &= \frac{1}{1 + e^{-a_1(N_c - c_1)}} \quad a_1 = -\frac{log(-(1-\theta)/\theta)}{N_c^* - c_1} \\ \delta &= \frac{1}{1 + e^{a_2(N_c - c_2)}} \quad a_2 = \frac{log(-(1-\theta)/\theta)}{N_c^* - c_2} \end{aligned} \tag{A.2}$$

where N_c is the total number of cells, N_c^* is the carrying capacity, θ is the turnover rate and $c_1, c_2 \in \mathbb{R}^+$. As for the linear case, the birth and rate functions in (A.1) and (A.2) satisfy $\beta(N_c^*) = \delta(N_c^*) = \theta$. The pie charts in Figure A.2 illustrate the distribution of filament lengths when reaching carrying capacity obtained with β and δ as in (A.1) and (A.2). The plot shows that in both cases, the fraction of long filaments in the population (i.e. with more than 2 cells) increases as the turnover decreases. However, with sigmoidal functions, the trend is not very marked. On the other hand, the charts obtained with hyperbolic functions are qualitatively comparable with the ones of Figure 1d of the main text. This difference is due to the fact that the sigmoidal birth rate drops quickly towards the turnover value and does not allow a remarkable elongation of the filaments. The hyperbolic birth rate function instead maintains high values for a longer time, before starting to decrease.

A.3 Length distribution during the transient phase

Figure A.4 show the mean length of the filaments calculated every iteration for two more values of θ in addition to $\theta = 0.9, 0.1$ presented in the main text. In Figure A.4, for the case of turnover $\theta = 0.01$ the mean filament length does not exceed 1000 cells. For a turnover one order of magnitude lower ($\theta = 0.001$), the peak of the average length of filaments (up to 2000 cells) is significantly higher. Statistical support is given in table A.2.

A.4 Comparison of the theoretical and experimental distributions

Fig. SA.6 shows the distribution of filament lengths obtained from experimental (*Fibrella aestuarina*) and simulation data at different time points. The histograms indicate the percentage of filaments in a given class of length, in terms of number of cells. We used an approximation of 6.7 μm of the cell size as we have indicated in § 2.3 of the main text. There is a 1-to-1 correspondence between number of cells and bins. Hence, the first bin of the histograms indicates the percentage of single

celled filaments, the second indicates the percentage of filaments with two cells, and so on. In order to make an illustrative comparison between the two datasets, we synchronize the two time-series according to the respective time or iteration at which they reach the carrying capacity. According to the OD curve, *F. aestuarina* reaches its carrying capacity after around 45 hours. For the theoretical data, we simulate the population dynamics for $\theta = 0.1$, using a starting conditions that that is comparable to the distribution of the experimental inoculum. With this initial condition, the population reaches the carrying capacity after 9 iterations. We have then

$$45 \text{ h} \simeq 9 \text{ iterations} \Rightarrow 5 \text{ h} \simeq 1 \text{ iteration}\,.$$

We considered the bell-shaped curve of the mean filament length vs time (Figure 2b in the main text) as a reference to choose the following five time points for comparison: the starting condition, one point at the left of the peak, the peak of the average length, one point at the right of the peak and the carrying capacity. The simulation-based histograms in Figure A.6 show a qualitative match to trends in experimental data. A precise quantitative comparison is not possible, as we do not have an estimation of the turnover of the *F. aestuarina* strain against time. In the beginning, each population is mainly made up of short filaments. As the mean length reaches its peak, the population has filaments of various lengths. After the peak, the distribution of lengths is again shifted to the left, whereby short filaments are more abundant than the long ones.

A.5 Mean length at carrying capacity and at the end of the experiment

Table A.1 provides the mean filament length measurements when the cultures reach the carrying capacity and at the end of the experiment. The number of cells was calculated according to the same conversion used in the main text (see § 2.3 of the main text). Heterotrophic bacteria have a fast turnover, hence according to the model prediction (see Figure 1c in the main text), their mean filament length

should be already close the one of the stationary state (two cells) when reaching carrying capacity. Data in Table A.1 support this hypothesis: the average number of cells per filament of all heterotrophic species is near two when reaching carrying capacity. Cyanobacteria exhibit a much slower generation time and hence a lower turnover. When reaching carrying capacity the filament length distribution is still far away from the stationary distribution of two cells per filament. Table A.1 confirms this prediction, showing that at that point cyanobacteria are still very long.

The experiments performed in this study were not run in a chemostat. Therefore, after the culture reached carrying capacity, nutrient availability was decreasing and consequently death rate was increasing. The breakage process has been accelerated and the carrying capacity of the population gradually dropped below the initial one. This observation explains why at the end of the heterotroph experiments, the average filament length is below two.

A.6 Statistical tests

For the simulation data, significance of the increase and decrease of mean filament length at different time points was proved by comparing the distribution of the mean at three reference points, namely the starting condition, the peak of the filament length and the time at which the population reaches its carrying capacity. Since the data were not normally distributed, we used a Wilcoxon rank sum test as a nonparametric hypothesis test. For each tested pair and for every turnover rate, the null hypothesis that the two compared samples come from distributions with equal medians was rejected, as shown in table A.2. The rank sum test was performed with MATLAB statistics toolbox (www.mathworks.ch/products/statistics/).

For the experimental data, significance of the increase and decrease of filament length at different time points was proved by a one-way analysis of variance (ANOVA, significance level of 0.05) followed by a Tukey's test as a multiple comparison procedure. For each bacterial species, data at different time points were considered as independent samples containing mutually independent observations. In all cases, in order to match the requirement of normality of the distribution, we transformed the sampled data by taking their natural logarithm. The ANOVA

test compares the sample means and returns the p-value under the null hypothesis that all samples are drawn from populations with the same mean. However, ANOVA evaluates the hypothesis that the samples all have the same mean against the alternative that the means are not all the same. In order to determine which pairs of means were significantly different, we performed a multiple comparison using the Tukey's honestly significant difference criterion. The output of the latter test for each species is shown in tables A.3-A.7. In the tables, for each time point pair we show the estimated difference in means and a confidence interval for this difference. If the confidence interval does not contain 0.0, the samples means are significantly different at the 0.05 level. If the confidence interval contains 0.0, than the the samples means are not significantly different at the 0.05 level. Bolded intervals correspond to the comparison between the inoculum and peak and between peak and carrying capacity. These three points were taken as a reference to prove significant increase/decrease of mean length. Both ANOVA and multiple comparison have been performed with MATLAB statistics toolbox. According to the ANOVA test, the hypothesis that the means are all the same was rejected in the case of all species ($p < 0.05$).

Statistical support for different distributions of filament length when reaching carrying capacity is shown in Tables A.8 and A.9, for the simulation and the experimental data respectively. In the case of the simulations, we compared the distribution for different turnovers at that time point. The datasets of the means of 1000 runs were normally distributed. Hence we applied a one-way ANOVA followed by a multiple comparison using the Tukey's honestly significant difference criterion. In the case of experiments, we compared the distribution of different species when the population size reached its carrying capacity. We took the same approach of a one-way ANOVA coupled with a multiple comparison, after a logarithmic transformation of the data to match the assumption of normality.

A.7 Serial Transfer experiment and simulation

Simulation. We run a series of successive simulations, each one consisting of 20 iterations (Fig. 5a in the main text). The initial curve was obtained with the default starting condition of one cell. When the average filament length of the

initial culture was at its maximum, a simulation of a new culture was started. The new simulation was initialized with one filament of the length corresponding to the average length at the peak of the previous culture (averaged over 50 runs). The same steps were repeated for each newly started culture. Each new simulation with low population density emulates a transfer to a fresh medium. The choice of a carrying capacity higher than the default of 5000 cells would have allowed the setting of a distribution of filaments instead of a single filament as a starting condition. However, increasing the carrying capacity leads to a considerably longer computational time. Thereby we opted for an approximation using the mean filament length as starting condition for each successive transfer.

Laboratory experiment. Bacteria were cultivated the same way as in the batch culture experiments. After 17 hours the first transfer was done, whereby five mL of the growing culture were transferred into 50-mL flasks containing 20 mL of R_2A medium. The initial culture was incubated on a shaker (120 rpm) at 29 ℃ for 5 hours, and then transferred again. Subsequently the new transfers were done with 3 hours interval. Before each transfer, sample aliquots were fixed with formaldehyde (final concentration, 2%) and used for microscopic length observation. Pictures were taken with a digital camera (Color View, Soft Imaging System) connected to a light microscope (Olympus BX 51, Germany) using a 4x objective. At least 450 filaments were counted per time point.

A.8 Estimation of the time of first spilt

The assumptions of the model presented in the main text allow the derivation of the estimated iteration at which the mean filament length in the population starts the decreasing phase. Recalling the definition of the birth and death rate functions β and δ as $\beta(N_c) = -c_1 N_c + 1,\ c_1 = (1-\theta)/N_c^*$ and $\delta(N_c) = c_2 N_c,\ c_2 = \theta/N_c^*$, and setting $N_c = i$, the probability that the next event is a death is given by

$$\text{Prob(death)} = \frac{\delta}{\beta + \delta} = \frac{(\theta i)/N_c^*}{1 - [(1-\theta)i]/N_c^* + (\theta i)/N_c^*} \sim \frac{\theta i}{N_c^*}$$

where the last approximation holds if $i \ll N_c^*$. The probability that an event is a birth is hence $1 - \frac{\theta i}{N_c^*}$. Assuming one single cell as a starting condition, the

probability that the first m events are births is given by

$$\prod_{i=1}^{m}(1-\frac{\theta i}{N_c^*}) \sim e^{-\sum_{i=1}^{m}\frac{\theta i}{N_c}} = e^{-\frac{\theta\, m(m+1)}{2N_c}} \sim e^{-\frac{\theta\, m^2}{2N_c}}\,, \tag{A.3}$$

where for the first approximation we used the fact that for any $\varepsilon \ll 1$,

$$1-\varepsilon = \exp(\log(1-\varepsilon)) = \exp(-\varepsilon - \frac{1}{2}\varepsilon - \frac{1}{3}\varepsilon \dots) \sim \exp(-\varepsilon)\,.$$

One can then use eq. (A.3) to estimate the number of cells occurring before the first filament split. For example one can start by assigning the typical value of $1/e$ to the probability defined by (A.3). We have then

$$e^{-\frac{\theta\, m^2}{2N_c}} = 1/e \Leftrightarrow \frac{\theta\, m^2}{2N_c} = 1 \Leftrightarrow m = \sqrt{(2N_c^*)/\theta}\,.$$

If we suppose that at any of the first m events every cell divides, and considering that the initial filament is a single cell, we expect to have 2^t cells after t iterations and hence

$$2^t \approx m \Rightarrow t \approx \log_2 m\,. \tag{A.4}$$

In (A.4), t represents the approximate iteration of first filament breakage. The values t obtained by plugging in the values of N_c^* and θ used in the paper fit well with the results showed in Figure 2 of the main text and Figure 3. The time of first split coincides roughly with the iteration at which the average length of filaments reaches its peak. For $\theta = 0.9, 0.1, 0.01, 0.001$ we have $t = 6.7, 8.2, 10, 11.5$, observed in Figures 2 of the main text and S3 respectively. The case of a different initial condition, where there is more than one single cell, can be treated similarly and has been tested. By letting the product in (A.3) to start from a value $N_0 > 1$, we obtain $m_0 = \sqrt{(2N_c^*/N_0)/\theta}$. The initial condition used to produce Figure 5 has been chosen to test the validity of the latter formula for m_0. The obtained result fitted again the predicted iteration of first split given by $t_0 \approx \log_2 m_0 = 4.6$, as shown in Figure A.7.

A.9 Stationary distribution of filament length based on an analogous continuous time model

By considering the possible events (birth or death of a cell) in any time interval, one can define the following model for the evolution of the filamentation process in continuous time. Let $N_j(t)$, $j \geq 1$, denote the number of filaments of length j in the system at time t. Then the Markovian transition rates out of the state $\overrightarrow{N} := (N_1, N_2, \ldots)$ are given by

$$\overrightarrow{N} \;\to\; \overrightarrow{N} + \hat{\mathbf{e}}^{(j+1)} - \hat{\mathbf{e}}^{(j)} \qquad \text{at rate} \quad j\, b(\overrightarrow{N})\, N_j, \quad j \geq 1 \tag{A.5}$$

$$\overrightarrow{N} \;\to\; \overrightarrow{N} + 2\hat{\mathbf{e}}^{(j)} - \hat{\mathbf{e}}^{(2j+1)} \qquad \text{at rate} \quad d(\overrightarrow{N})\, N_{2j+1}, \quad j \geq 1 \tag{A.6}$$

$$\overrightarrow{N} \;\to\; \overrightarrow{N} + \hat{\mathbf{e}}^{(j-k-1)} + \hat{\mathbf{e}}^{(k)} - \hat{\mathbf{e}}^{(j)} \qquad \text{at rate} \quad 2\, d(\overrightarrow{N})\, N_j, \quad 0 \leq k < \frac{j-1}{2},\ j \geq 1 \tag{A.7}$$

where $\hat{\mathbf{e}}^{(j)}, j \geq 1$, denotes the j-th unit vector, and $b(\overrightarrow{N})$, $d(\overrightarrow{N})$ represent the *per capita* birth and death rates respectively. The latter are given by

$$b(\overrightarrow{N}) \;:=\; b - c_1 N_c / N_c^*, \qquad d(\overrightarrow{N}) \;:=\; c_2 N_c / N_c^*,$$

where $b = c_1 + c_2$, $N_c := \sum_{j\geq 1} j N_j$ is the total number of cells in the system, and N_c^* is the value of N_c at the carrying capacity. In equations (A.5)-(A.7), the first transition refers to cell birth in a filament of length j, whereby filaments of length j lose one unit ($-\hat{\mathbf{e}}^{(j)}$) and those of length $j+1$ increase by one ($+\hat{\mathbf{e}}^{(j+1)}$). The second transition refers to cell death at the center of a filament with an odd number of cells. In this case, a filament with $2j+1$ cells breaks ($-\hat{\mathbf{e}}^{(2j+1)}$) in two filaments of length j ($+2\hat{\mathbf{e}}^{(j)}$). The third transition represents the case of a cell death at any position of a filament other than the exact center, whereby a filament of length j breaks at position $k+1$ ($-\hat{\mathbf{e}}^{(j)}$) producing a filament of length $j-k-1$ ($+\hat{\mathbf{e}}^{(j-k-1)}$) and one of length k ($+\hat{\mathbf{e}}^{(k)}$).

In the long term, when there are many cells in the system, the ratios $n_j := N_j/N_c^*$ can be thought of as more or less continuous quantities, with behavior the

average of what is predicted by the transition rates in (A.5)-(A.6). This yields the following differential equations describing the evolution of the n_j:

$$\frac{dn_j}{dt} = (b - c_1 m)(j-1)n_{j-1} - j(b - (c_1 - c_2)m)n_j + 2c_2 m \sum_{i>j} n_i, \qquad \text{(A.8)}$$

where $m := \sum_{j\geq 1} j n_j$ represents the total number of cells as a proportion of the carrying capacity. Formally adding over $j \geq 1$, one obtains

$$\frac{dx}{dt} = c_2 m(m - 2x), \qquad \text{(A.9)}$$

where $x := \sum_{j\geq 1} n_j$ is the total number of filaments weighted by the carrying capacity. Then, recalling $c_1 + c_2 = b$,

$$\frac{dm}{dt} = b(1 - m)m. \qquad \text{(A.10)}$$

Equation (A.10) is the well-known logistic equation, and can be integrated to show exponentially fast convergence of $m(t)$ to its limit 1; it then follows by integrating (A.9) that $x(t) \to 1/2$. Considering that the average length of filaments $\overline{L}$ can be derived as

$$\overline{L} = \frac{m}{x} = \frac{\sum_{j\geq 1} j N_j / N_c^*}{\sum_{j\geq 1} N_j / N_c^*} = \frac{\sum_{j\geq 1} j N_j}{\sum_{j\geq 1} N_j}$$

and that, as mentioned above, $m(t) \to 1, x(t) \to 1/2$, we have that

$$\overline{L} \to \frac{1}{1/2} = 2.$$

In fact, more is true: the equations (A.8) have a stationary solution with $m = 1$ and $n_j = 2^{-j}$, $j \geq 1$, so that the equilibrium distribution of filament lengths is geometric with mean 2.

	Carrying capacity		Last datapoint	
Species	Mean length	Confidence interval	Mean length	Confidence interval
R. lutea A	3.1	[0.7 , 9.1]	1.5	[0.7 , 3.9]
R. lutea B	4.3	[0.9 , 13]	1.8	[0.6 , 5]
F. limi	2.7	[0.7 , 10.4]	1.4	[0.4 , 4.1]
F. aestuarina A	4.2	[0.6 , 18.8]	1	[0.3 , 2.8]
F. aestuarina B	3.8	[0.7 , 15.9]	1.2	[0.3 , 4.4]
N. muscorum A	87.9	[9 , 305.1]	74.7	[5.7 , 274.7]
N. muscorum B	102	[10.2 , 324.6]	74.4	[7.8 , 243.8]
A. variabilis A	166.3	[11.3 , 378.3]	30.7	[3.7 , 119.4]
A. variabilis B	160.3	[28.1 , 370]	25.3	[4 , 92.3]

Table A.1: Mean length of filaments expressed in number of cells when reaching carrying capacity and at the end of the experiment. The confidence interval is derived from the 25th and 75th percentiles of the measured lengths. Capital letters next to the species name indicate the experiment.

Turnover	Pair compared	p-value	Significant difference
$\theta = 0.9$	Starting point - Peak	0	yes
	Peak - carrying capacity	0	yes
$\theta = 0.1$	Starting point - Peak	0	yes
	Peak - carrying capacity	0	yes
$\theta = 0.01$	Starting point - Peak	0	yes
	Peak - carrying capacity	0	yes
$\theta = 0.001$	Starting point - Peak	0	yes
	Peak - carrying capacity	0	yes

Table A.2: Wilcoxon rank-sum test to prove significant increase and decrease in mean filament length in the simulation data, for different turnovers.

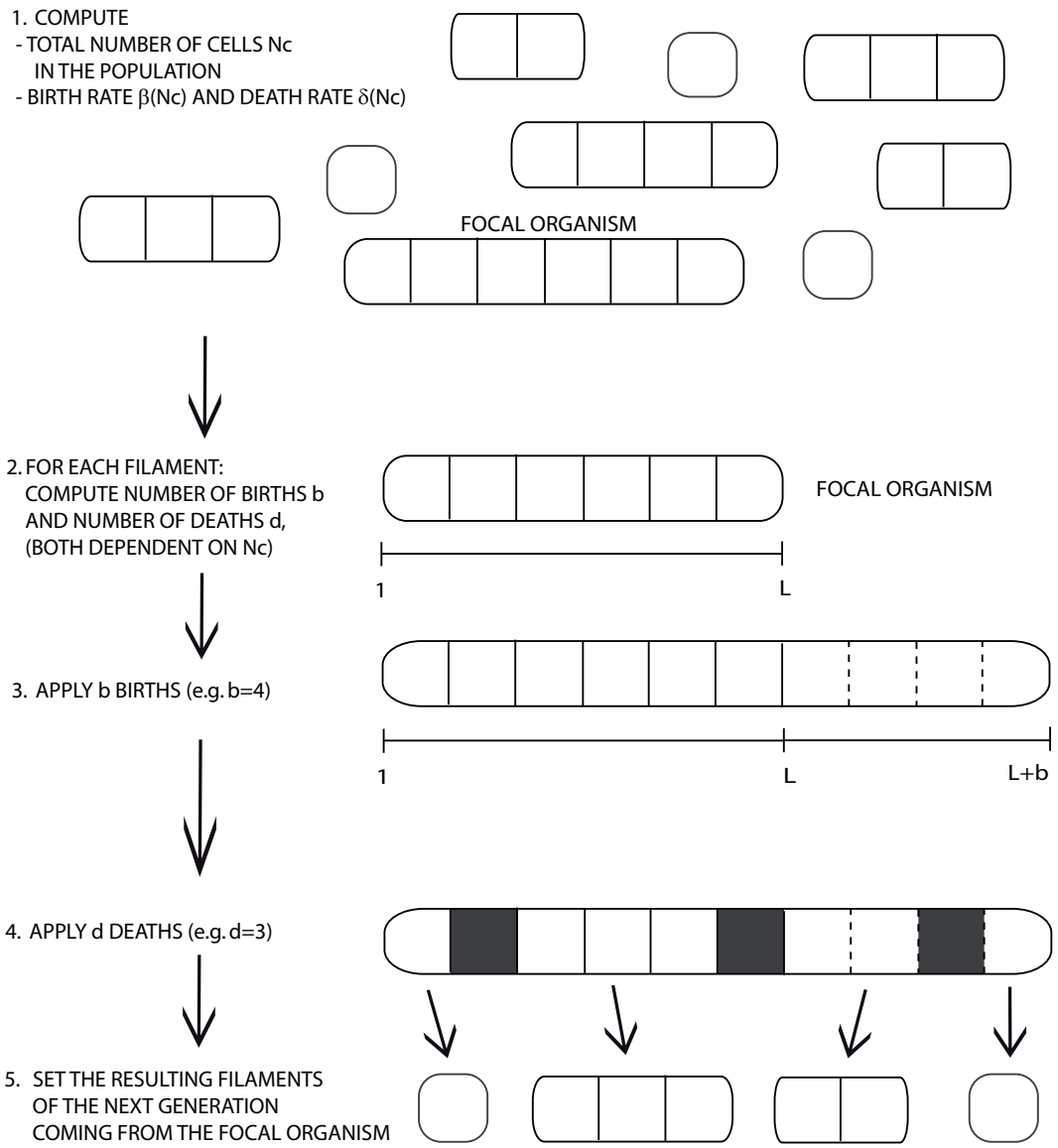


Figure A.1: Schematic view of the algorithm describing how filament length distributions are affected by processes governing birth and death rates of cells.

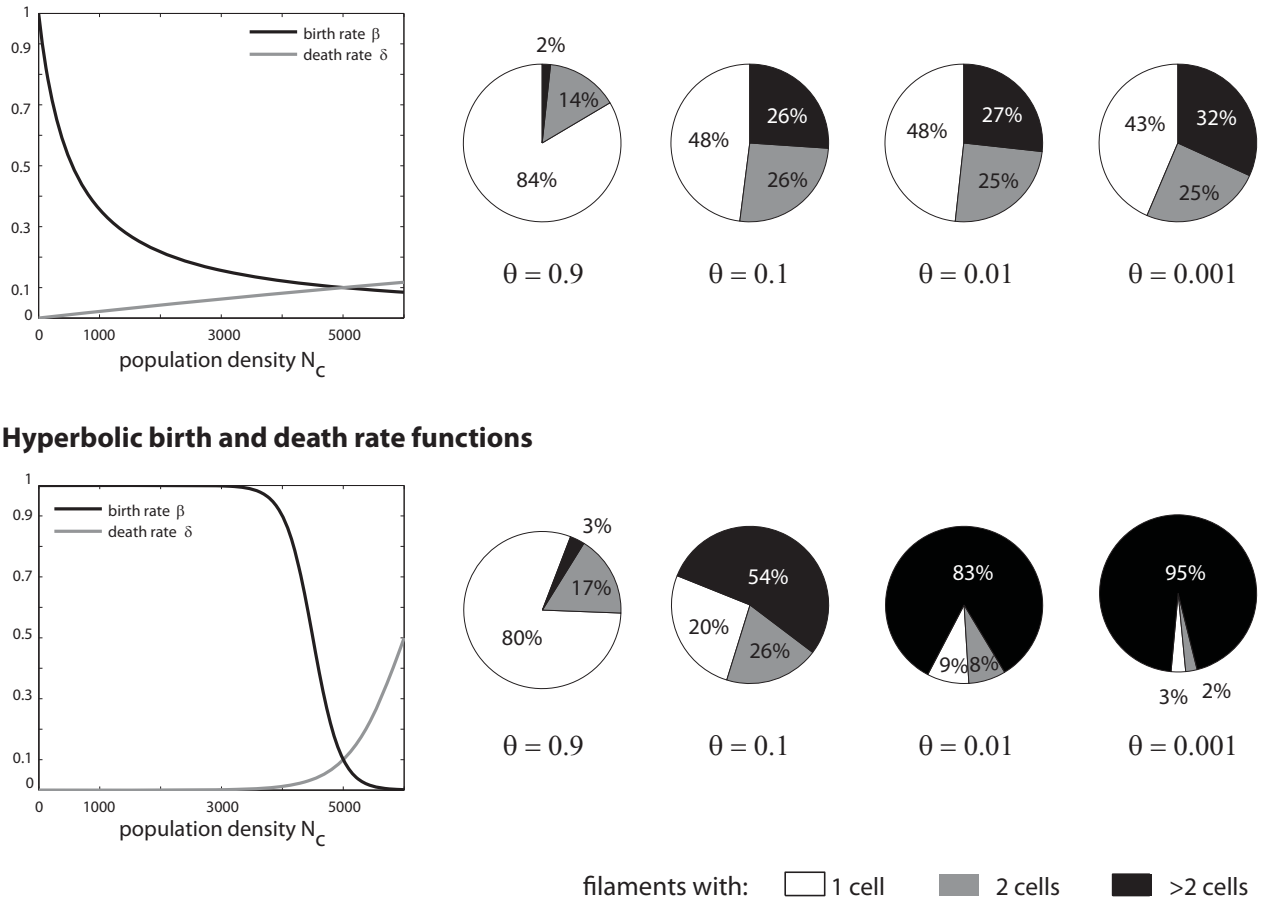


Figure A.2: Distribution of filament length when reaching carrying capacity at $\theta = 0.1$, for the case of nonlinear birth and death rates.

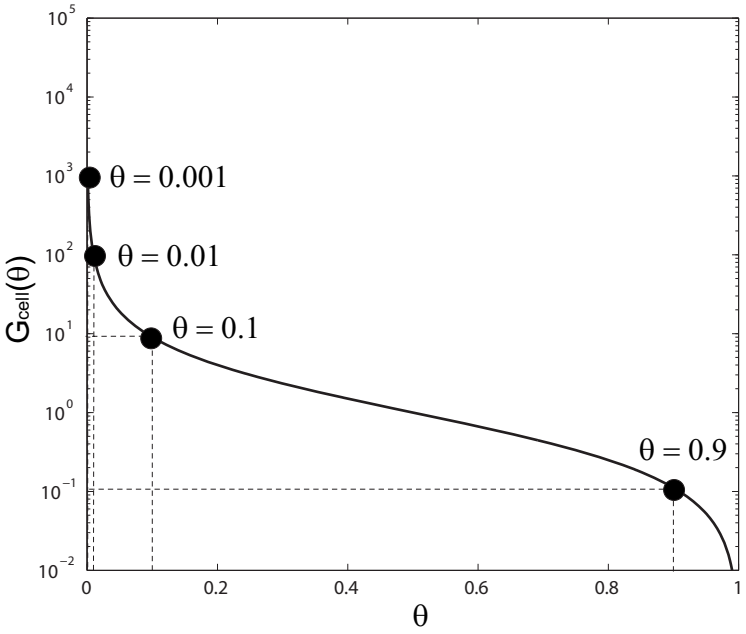


Figure A.3: Plot of the function of generation time$G_{cell}(\theta) = \theta \mapsto (1-\theta)/\theta$ for $\theta \in [0,1]$ (log scale). The generation time of a cell is a decreasing function of the turnover rate. The dots indicate the generation time $G_{cell}(\theta)$ at carrying capacity correspondent to the values of θ used in the simulations.

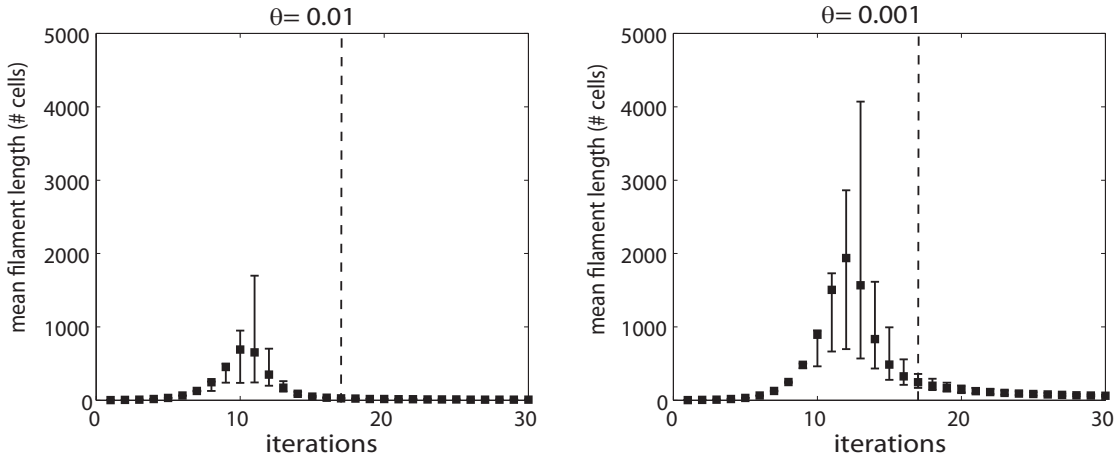


Figure A.4: The mean filament length for 30 iterations has been tracked for $\theta = 0.01$ and $\theta = 0.001$ and averaged over 1000 runs. Edges of the error bars correspond to the 2.5th and 97.5th percentiles of the distribution of the mean filament lengths in the 1000 runs. As already shown in the main text, in the transient phase filaments are shorter if the turnover is high (left) while they can be considerably longer for low turnovers (right). The dotted line indicates the iteration at which the carrying capacity is reached.

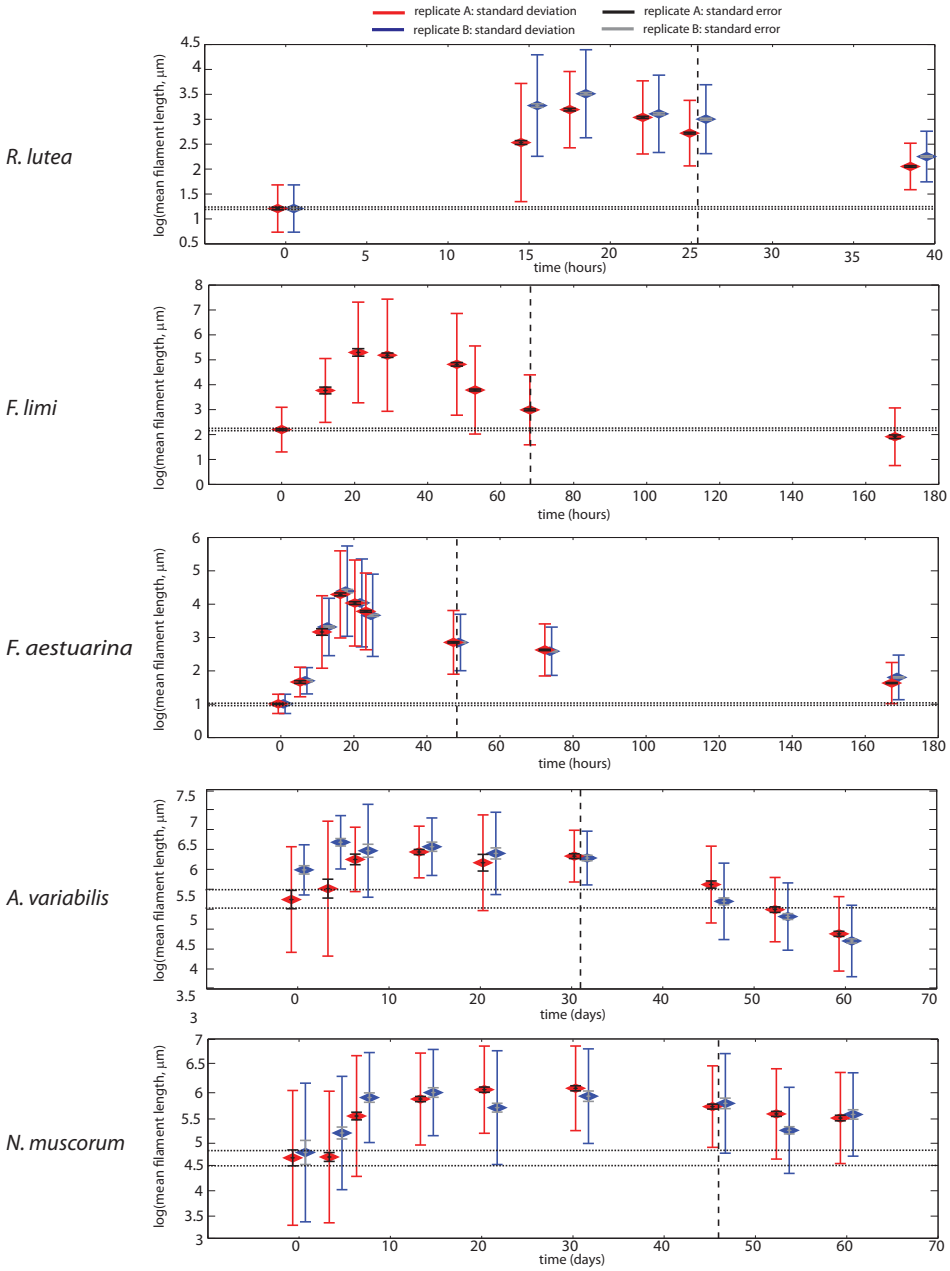


Figure A.5: Empirical measurements showing mean length of bacteria against time, after logarithmic transformation of the data. Standard deviation of the observation and standard error of the mean are shown. Diamonds indicate the logarithm of the mean length at each time point. The zero time point corresponds to the inoculum. The dotted vertical line indicates the time at which the carrying capacity was reached. The horizontal dotted lines indicate the interval determined by the standard error of the inoculum. It helps the reader in identifying the time points at which the mean length falls in the same range. The lines are indicated only for replicate A.

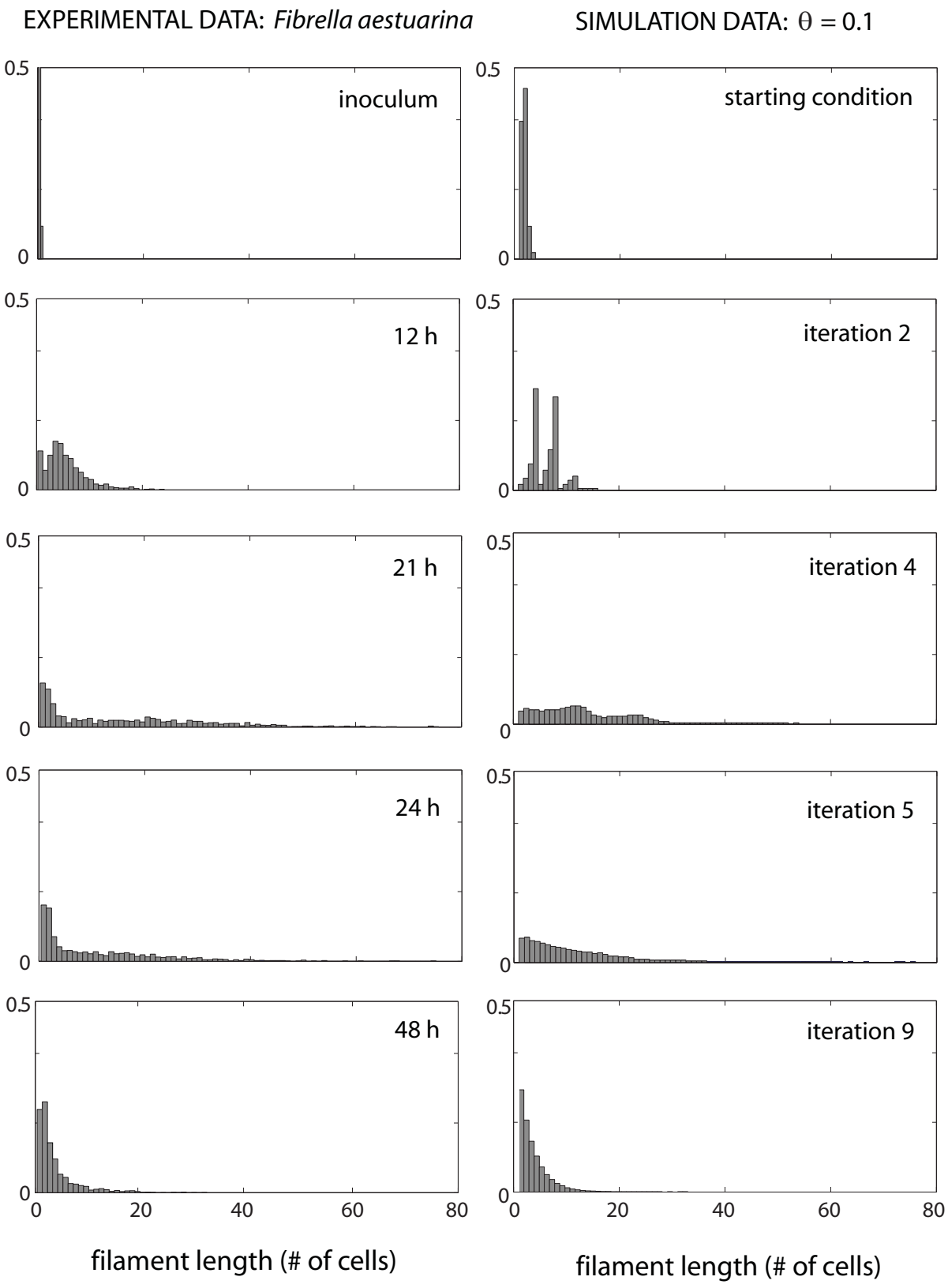


Figure A.6: Histograms of the distribution of filament lengths according to experimental data (left) and simulations (right).

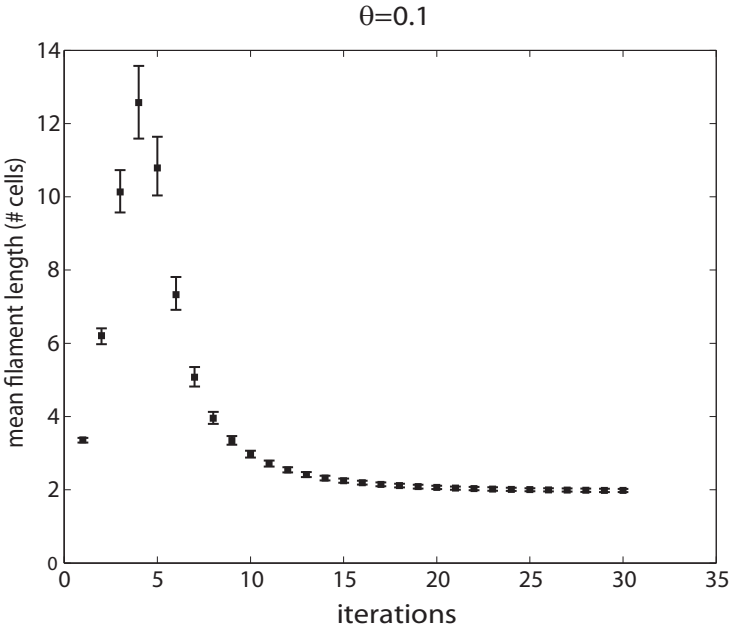


Figure A.7: Mean filament length over 1000 runs when the starting condition for the simulation is comparable to the distribution of the experimental inoculum of *Fibrella aestuarina*. Edges of the error bars correspond to the 2.5th and 97.5th percentiles of the computed means.

Time 1	Time 2	estimated difference in means	95% confidence interval of the difference	Significant difference
1	2	-1.3245	[-1.4587 -1.1903]	yes
1	3	**-1.9843**	**[-2.1198 -1.8488]**	**yes**
1	4	-1.8289	[-1.9572 -1.7007]	yes
1	5	-1.5131	[-1.6408 -1.3854]	yes
1	6	-0.8446	[-0.9781 -0.7111]	yes
2	3	-0.6598	[-0.7624 -0.5572]	yes
2	4	-0.5044	[-0.5973 -0.4116]	yes
2	5	-0.1886	[-0.2807 -0.0965]	yes
2	6	0.4799	[0.3799 0.5799]	yes
3	4	0.1554	[0.0607 0.2500]	yes
3	5	**0.4712**	**[0.3773 0.5651]**	**yes**
3	6	1.1397	[1.0380 1.2414]	yes
4	5	0.3158	[0.2327 0.3990]	yes
4	6	0.9843	[0.8925 1.0762]	yes
5	6	0.6685	[0.5774 0.7596]	yes

Table A.3: Multiple comparison for *R. lutea*. The numbers in bold pertain to comparisons between the time points taken as a reference to prove the increase or decrease of filament length (between inoculum, peak and carrying capacity). Time points 1, 3 and 5 correspond to the inoculum, peak and carrying capacity, respectively.

Time 1	Time 2	estimated difference in means	95% confidence interval of the difference	Significant difference
1	2	-0.7856	[-1.0727 -0.4984]	yes
1	**3**	**-1.5483**	[**-1.7629 -1.3337**]	**yes**
1	4	-1.4924	[-1.6317 -1.3531]	yes
1	5	-1.3098	[-1.4425 -1.1772]	yes
1	6	-0.7944	[-0.9058 -0.6830]	yes
1	7	-0.3966	[-0.5153 -0.2779]	yes
1	8	0.1428	[-0.0432 0.3288]	no
2	3	-0.7627	[-1.0944 -0.4311]	yes
2	4	-0.7068	[-0.9955 -0.4181]	yes
2	5	-0.5243	[-0.8098 -0.2387]	yes
2	6	-0.0088	[-0.2851 0.2675]	no
2	7	0.3889	[0.1096 0.6683]	yes
2	8	0.9284	[0.6145 1.2423]	yes
3	4	0.0559	[-0.1608 0.2726]	no
3	5	0.2385	[0.0260 0.4509]	yes
3	6	0.7539	[0.5540 0.9538]	yes
3	**7**	**1.1517**	[**0.9476 1.3557**]	**yes**
3	8	1.6911	[1.4418 1.9404]	yes
4	5	0.1826	[0.0465 0.3186]	yes
4	6	0.6980	[0.5826 0.8134]	yes
4	7	1.0958	[0.9733 1.2183]	yes
4	8	1.6352	[1.4468 1.8236]	yes
5	6	0.5154	[0.4082 0.6227]	yes
5	7	0.9132	[0.7984 1.0280]	yes
5	8	1.4526	[1.2691 1.6362]	yes
6	7	0.3978	[0.3084 0.4871]	yes
6	8	0.9372	[0.7684 1.1060]	yes
7	8	0.5394	[0.3657 0.7131]	yes

Table A.4: Multiple comparison for *F. limi*. Time points 1,3 and 7 correspond to the inoculum, peak and carrying capacity respectively.

Time 1	Time 2	estimated difference in means	95% confidence interval of the difference	Significant difference
1	2	-0.6569	[-0.9532 -0.3606]	yes
1	3	-2.1569	[-2.3096 -2.0042]	yes
1	4	**-3.2831**	**[-3.4351 -3.1311]**	**yes**
1	5	-3.0246	[-3.1738 -2.8753]	yes
1	6	-2.7745	[-2.9046 -2.6443]	yes
1	7	-1.8450	[-1.9741 -1.7159]	yes
1	8	-1.6181	[-1.7553 -1.4810]	yes
1	9	-0.6248	[-0.7617 -0.4879]	yes
2	3	-1.5000	[-1.7937 -1.2063]	yes
2	4	-2.6262	[-2.9195 -2.3329]	yes
2	5	-2.3677	[-2.6595 -2.0758]	yes
2	6	-2.1176	[-2.4002 -1.8350]	yes
2	7	-1.1881	[-1.4702 -0.9060]	yes
2	8	-0.9612	[-1.2471 -0.6753]	yes
2	9	0.0321	[-0.2537 0.3179]	no
3	4	-1.1262	[-1.2730 -0.9795]	yes
3	5	-0.8677	[-1.0116 -0.7238]	yes
3	6	-0.6176	[-0.7416 -0.4936]	yes
3	7	0.3119	[0.1890 0.4348]	yes
3	8	0.5388	[0.4074 0.6701]	yes
3	9	1.5321	[1.4010 1.6632]	yes
4	5	0.2585	[0.1154 0.4017]	yes
4	6	0.5086	[0.3856 0.6317]	yes
4	7	**1.4381**	**[1.3161 1.5601]**	**yes**
4	8	1.6650	[1.5345 1.7955]	yes
4	9	2.6583	[2.5281 2.7885]	yes
5	6	0.2501	[0.1304 0.3698]	yes
5	7	1.1796	[1.0610 1.2981]	yes
5	8	1.4064	[1.2792 1.5337]	yes
5	9	2.3998	[2.2727 2.5268]	yes
6	7	0.9295	[0.8361 1.0228]	yes
6	8	1.1563	[1.0521 1.2606]	yes
6	9	2.1497	[2.0458 2.2536]	yes
7	8	0.2269	[0.1240 0.3298]	yes
7	9	1.2202	[1.1176 1.3228]	yes
8	9	0.9933	[0.8808 1.1059]	yes

Table A.5: Multiple comparison for *F. aestuarina*. Time points 1, 4 and 7 correspond to the inoculum, peak and carrying capacity respectively.

Time 1	Time 2	estimated difference in means	95% confidence interval of the difference	Significant difference
1	2	-0.2734	[-0.9283 0.3815]	no
1	3	-1.0062	[-1.7081 -0.3043]	yes
1	4	**-1.1920**	**[-1.7798 -0.6042]**	**yes**
1	5	-0.9229	[-1.6446 -0.2013]	yes
1	6	-1.0886	[-1.6670 -0.5103]	yes
1	7	-0.3759	[-0.9548 0.2030]	no
1	8	0.2530	[-0.3166 0.8227]	no
1	9	0.8612	[0.3136 1.4087]	yes
2	3	-0.7328	[-1.3658 -0.0997]	yes
2	4	-0.9185	[-1.4221 -0.4150]	yes
2	5	-0.6495	[-1.3044 0.0054]	no
2	6	-0.8152	[-1.3077 -0.3227]	yes
2	7	-0.1025	[-0.5956 0.3907]	no
2	8	0.5265	[0.0441 1.0088]	yes
2	9	1.1346	[0.6786 1.5906]	yes
3	4	-0.1858	[-0.7491 0.3775]	no
3	5	0.0832	[-0.6187 0.7851]	no
3	6	-0.0824	[-0.6359 0.4710]	no
3	7	0.6303	[0.0762 1.1844]	yes
3	8	1.2592	[0.7148 1.8036]	yes
3	9	1.8673	[1.3461 2.3886]	yes
4	5	0.2690	[-0.3188 0.8568]	no
4	6	**0.1033**	**[-0.2956 0.5023]**	**no**
4	7	0.8161	[0.4163 1.2158]	yes
4	8	1.4450	[1.0587 1.8313]	yes
4	9	2.0531	[1.7003 2.4060]	yes
5	6	-0.1657	[-0.7440 0.4127]	no
5	7	0.5470	[-0.0319 1.1260]	no
5	8	1.1760	[0.6063 1.7457]	yes
5	9	1.7841	[1.2365 2.3317]	yes
6	7	0.7127	[0.3270 1.0985]	yes
6	8	1.3417	[0.9699 1.7134]	yes
6	9	1.9498	[1.6129 2.2867]	yes
7	8	0.6289	[0.2563 1.0016]	yes
7	9	1.2371	[0.8992 1.5749]	yes
8	9	0.6081	[0.2863 0.9299]	yes

Table A.6: Multiple comparison for *A. variabilis*. Time points 1,4 and 6 correspond to the inoculum, peak and carrying capacity respectively.

Time 1	Time 2	estimated difference in means	95% confidence interval of the difference	Significant difference
1	2	-0.0195	[-0.4886 0.4496]	no
1	3	-0.9441	[-1.4060 -0.4821]	yes
1	4	-1.3273	[-1.7791 -0.8754]	yes
1	5	-1.5394	[-1.9966 -1.0821]	yes
1	6	**-1.5702**	**[-2.0264 -1.1140]**	yes
1	7	-1.1567	[-1.6141 -0.6993]	yes
1	8	-0.9917	[-1.4422 -0.5413]	yes
1	9	-0.8992	[-1.3577 -0.4408]	yes
2	3	-0.9246	[-1.2437 -0.6055]	yes
2	4	-1.3078	[-1.6122 -1.0034]	yes
2	5	-1.5199	[-1.8322 -1.2076]	yes
2	6	-1.5507	[-1.8615 -1.2399]	yes
2	7	-1.1372	[-1.4497 -0.8247]	yes
2	8	-0.9723	[-1.2745 -0.6700]	yes
2	9	-0.8798	[-1.1938 -0.5657]	yes
3	4	-0.3832	[-0.6764 -0.0900]	yes
3	5	-0.5953	[-0.8967 -0.2940]	yes
3	6	-0.6262	[-0.9260 -0.3263]	yes
3	7	-0.2126	[-0.5142 0.0890]	no
3	8	-0.0477	[-0.3386 0.2432]	no
3	9	0.0448	[-0.2584 0.3480]	no
4	5	-0.2121	[-0.4978 0.0736]	no
4	6	-0.2430	[-0.5271 0.0412]	no
4	7	0.1706	[-0.1154 0.4566]	no
4	8	0.3355	[0.0608 0.6102]	yes
4	9	0.4280	[0.1403 0.7157]	yes
5	6	-0.0308	[-0.3234 0.2617]	no
5	7	0.3827	[0.0883 0.6771]	yes
5	8	0.5476	[0.2642 0.8311]	yes
5	9	0.6401	[0.3441 0.9362]	yes
6	7	**0.4135**	**[0.1207 0.7064]**	yes
6	8	0.5785	[0.2967 0.8603]	yes
6	9	0.6710	[0.3765 0.9654]	yes
7	8	0.1649	[-0.1188 0.4486]	no
7	9	0.2574	[-0.0389 0.5537]	no
8	9	0.0925	[-0.1929 0.3779]	no

Table A.7: Multiple comparison for *N. muscorum*. Time points 1,6 and 7 correspond to the inoculum, peak and carrying capacity respectively.

Turnover 1	Turnover 2	estimated difference in means	95% confidence interval of the difference	Significant difference
0.001	0.01	222.7728	[219.4576 226.0879]	yes
0.001	0.1	244.8196	[241.5044 248.1347]	yes
0.001	0.9	247.1738	[243.8587 250.4890]	yes
0.01	0.1	22.0468	[18.7316 25.3620]	yes
0.01	0.9	24.4011	[21.0859 27.7162]	yes
0.1	0.9	2.3542	[-0.9609 5.6694]	no

Table A.8: Multiple comparison for the distribution of filament lengths when reaching carrying capacity for different turnovers.

Species 1	Species 2	estimated difference in means	95% confidence interval of the difference	Significant difference
R. lutea	*F. limi*	0.2258	[0.1411 0.3104]	yes
R. lutea	*F. aestuarina*	-0.1325	[-0.2086 -0.0564]	yes
R. lutea	*N. muscorum*	-2.7501	[-2.8968 -2.6033]	yes
R. lutea	*A. variabilis*	-3.6092	[-3.8242 -3.3942]	yes
F. limi	*F. aestuarina*	-0.3583	[-0.4350 -0.2815]	yes
F. limi	*N. muscorum*	-2.9758	[-3.1230 -2.8287]	yes
F. limi	*A. variabilis*	-3.8350	[-4.0502 -3.6197]	yes
F. aestuarina	*N. muscorum*	-2.6176	[-2.7599 -2.4752]	yes
F. aestuarina	*N. muscorum*	-3.4767	[-3.6887 -3.2647]	yes
N. muscorum	*A. variabilis*	-0.8591	[-1.1055 -0.6128]	yes

Table A.9: Multiple comparison within species of the distribution of filament length when the population reaches its carrying capacity.

B Supplementary material of Chpt. 3

B.1 Basic Results without Cheaters

We depict how the resident population would grow without the presence of the cheaters. We recall that the system was found to evolve towards one of the following three steady states: one corresponding to the extinction of resident and cheater populations (Y_1), one in which only resident population survives (Y_2), and one in which the cheaters overcome the resident population (Y_3). Assuming a given density of vegetative cells and no heterocysts at time $t = 0$, at the beginning the most relevant role is played by nitrate, since the initial presence of some nitrate is necessary to allow any development and growth of cells. In fact, nitrate is only produced by heterocysts, but these can appear in the system only by division of vegetative cells. In the absence of nitrate, the growth of vegetatives is impeded, consequently the system can only evolve toward steady state Y_1 (Figure B.1aa-b). The same argument is not valid for solar energy, because irradiance represents a constant positive inflow rate (Figure B.1ac-d).

An extreme condition that also does not allow the growth of the resident population is either $p_v = 0$ or $p_h = 0$ (Figure B.1b). In the first case, the vegetative cells generate only heterocyst cells: they will both become extinct, because the vegetatives do not reproduce after their extinction, and hence heterocysts can not be produced anymore. If $p_h = 0$, no heterocysts and consequently no fixed nitrogen are produced and thus the vegetatives don't have the necessary resources to survive.

If these extreme cases are avoided and the death rate is sufficiently small to allow for growth, then in the absence of cheaters, the resident population follows a logistic trend, eventually reaching a steady state Y_2 (Figure B.1c).

B.2 Montecarlo Simulations for winning factor determination

Figure 4 of the main text shows the output of simulations in which only $p_v, p_{v'}$ have been chosen in $[0, 1]$, while the other constants are set according to Table 1 in the main text. Those results correspond to a Montecarlo sampling of 115,000 different parameter sets. However, plots of samples of size 15,000 are shown, because of better graphic appearance and consequent better comprehension of their related meaning.

To expand the domain of tested parameter sets, we used a Montecarlo sampling scheme in the entire parameter and initial conditions space, for both models.

In the homogeneous mix model, we sampled 115000 different sets of 21 values: the constant parameters a,q,r,c_0,I,r_0,k_0,k_N,k_C,k_{NC},k_I,k_G,k_{IG} in $[0, 100]$, k in $[0, 10]$, the initial densities $V(0), H(0), V'(0)$ in the range $[0, 100]$ and $N(0), C(0)$ in $[0, 10]$. We decided not to sample the death/decay rate p_3, because a slightly higher value of this parameter in comparison to others would have simply caused the death of the populations, giving then useless information about the investigated dynamics. In Figure B.2a we plot the vegetative steady states vs the ratio $p_v/p_{v'}$. Blue dots mean $V > V'$ and correspond to strains with p_v, red dots indicate that $V' > V$ and they compete having a value $p_{v'}$. The blue dots occupy the region of the graph where $p_v > p_{v'}$, and the red dots occupy the region where $p_{v'} > p_v$. Hence again the strain with the higher p_v value wins.

In the compartmental model we sampled sets of 22 values: the constant parameters as for the homogeneous mix model and the initial densities $V(0), H(0)$ in the range $[0, 100]$ and $N(0), C(0), N'(0), C'(0)$ in $[0, 10]$. We set equal initial values of vegetative and heterocyst cells for the two compartments, namely $V'(0) = V(0), H'(0) = H(0)$. In Figure B.2b, we plot the vegetative steady states after 115000 competitions. The legend of the dots is the same as in the previous figures. The horizontal axis is the ratio v_*/v'_*, where v_* indicates the steady state reached in isolation with the same correspondent parameters. The demarcation line around $v_*/v'_* = 1$ shows that the strain with the higher carrying capacity in isolation wins. The exceptions to the latter rule represent the 2%-3% of the competition experiments. These exceptions correspond to cases in which (i) a nu-

merical error occurs; (ii) $p_v \simeq p_{v'}$, hence the time required for the populations to stabilize is longer than the simulation time; (iii) The p_v value of the loser is too close to 0, hence the corresponding population can not grow.

B.3 Evolutionary optimization of vegetative/heterocyst ratio

To test the evolutionary optimization of p_v values, we simulated successive competitions by running stochastic simulations of the population dynamics through G_{max} mutational events. This means that we evaluate G_{max} consecutive competition experiments between a resident strain and a newly arrived mutant. We refer to the fixed population as the wild type and to the other competing population as the mutant, with the proportion of vegetative cells produced at each division equal to p_{wild} and p_{mutant} respectively. The simulation of the i mutant competitions, $i = 1 \dots G_{max}$, given an initial wild type with $p_{wild}(0)$ chosen randomly in $[0, 1]$, consists of the following steps:

1. Compute $p_{mutant}(i)$, using $p_{mutant}(i) = p_{wild}(i) + \delta_p(i)$, where $\delta_p \sim N(0, 1)$. If $p_{mutant}(i) > 1$ then we set $p_{mutant}(i) = 1$, if $p_{mutant}(i) < 0$ then $p_{mutant}(i) = 0$ (in order to assure $p_{mutant}(i) \in [0, 1]$).

2. Run the competition between the two strains, with equal numbers of wild type and mutant cells as initial condition.

3. Record the value of p_v for the winner and repeat step 2 with $p_{wild}(i + 1) = p_{wild}(i)$ if the wild type won the competition i, otherwise with $p_{wild}(i + 1) = p_{mutant}(i)$ if the mutant won the competition i.

We repeat the simulation for M trials, hence recording evolution along G_{max} mutant competitions with M different p_{wild} as starting point. We also ran simulations setting the number of cells for the wild type initial condition of generation i at the final steady state of the winner population at generation $i - 1$, but this was without effect on the final outcome.

B.4 Model of the Circadian Rhythm

In order to model the alternation of day and night in the circadian rhythm, we express the irradiance I by a Hill transformed sine curve as plotted in Figure B.3. We have

$$I(t) = A\frac{(\rho(t)+1)^\gamma}{(m^\gamma + (\rho(t)+1)^\gamma)}, \qquad \rho(t) = \sin(\frac{\pi t}{12})\,.$$

The parameter m regulates the duration of day and light. In particular, m is the value of $\rho(t)$ at which $I(t) = A/2$. We map the values of m into a quantity representing the percentage of daylight in respect to the 24 hours period. To achieve such mapping, we consider the values of t at which $I(t) = A/2$ and we take the difference between those values as an approximation of duration of day and night. Let t_1, t_2 be the first two points at which $I(t) = A/2$ (Figure B.3). Setting d as the percentage of day, we have that

$$\begin{aligned} t_1 &= \frac{12}{\pi} - \arcsin(m-1), \\ t_2 &= 12 - \frac{12}{\pi} - \arcsin(m-1) \\ d = \%\ \text{day} &= \frac{t_2 - t_1}{24} = \frac{12}{\pi}(\pi - 2\arcsin(m-1)) \\ \Rightarrow m &= 1 - \sin(d\pi - \frac{\pi}{2}) \end{aligned} \tag{B.1}$$

The case of no daylight corresponds to $d = 0, m = 2$. However, for $m = 2$ we don't have $I(t) = 0\ \ \forall t$, because I is a smooth function. Hence, to investigate this extreme case, we simply set $I = 0$ in the equations.

B.5 Phylogeny of Cyanobacteria

Sequence alignments of the 16S rRNA gene sequences (38 taxa, 2017 characters) were carried out using Clustal X software with default settings [200]. Phylogenetic reconstruction was carried out using a maximum likelihood analysis and Bayesian analyis. The best evolutionary model of nucleotide substitution that fits the data was obtained by using the Akaike Information Criterion as implemented in Mod-

eltest 3.5 [143]. The selected model was GTR + I + G (General time reversibel model, I: proportion of invariable sites, G: γ correction) [103, 154]. The estimated substitution rates were $R_{(A-C)}$ = 1.0489, $R_{(A-G)}$ =2.3685, $R_{(A-T)}$ =1.5135, $R_{(C-G)}$=0.4953, $R_{(C-T)}$=3.4634, $R_{(G-T)}$=1.0000. The base frequencies were estimated at 0.2479(A), 0.2159(C), 0.3071(G) and 0.2290(T). The proportion of invariable sites was estimated to be 0.3705 and the rate of heterogeneity among variable sites was estimated to follow a gamma distribution with the shape parameter α = 0.4380. Maximum likelihood analysis with the GTR+I+G model was performed using PAUP 4.0 [197]; bootstrap values were obtained from 400 resamplings of the data set. Bayesian analyis used a GTR+I+G model as well, but in this case, specific values of model parameters were not defined *a priori*. The analysis was done by running a Markov Chain Monte Carlo search with four chains, one cold and three heated, for 1,500,000 generations, with trees being sampled every 100 generations. To determine the "burn-in", log likelihood plots were examined for stationarity (where a plateau is reached). Stationarity was clearly reached after less than 10,000 generations but we discarded the first 3,000 trees to ensure that stationarity was completely reached. A higher "burn-in" did not alter the tree topology. Bayesian posterior probabilities were given by the percentage each branch was produced and were calculated from the remaining 12,001 trees. Bayesian analyis was performed using MrBayes 3 [157].

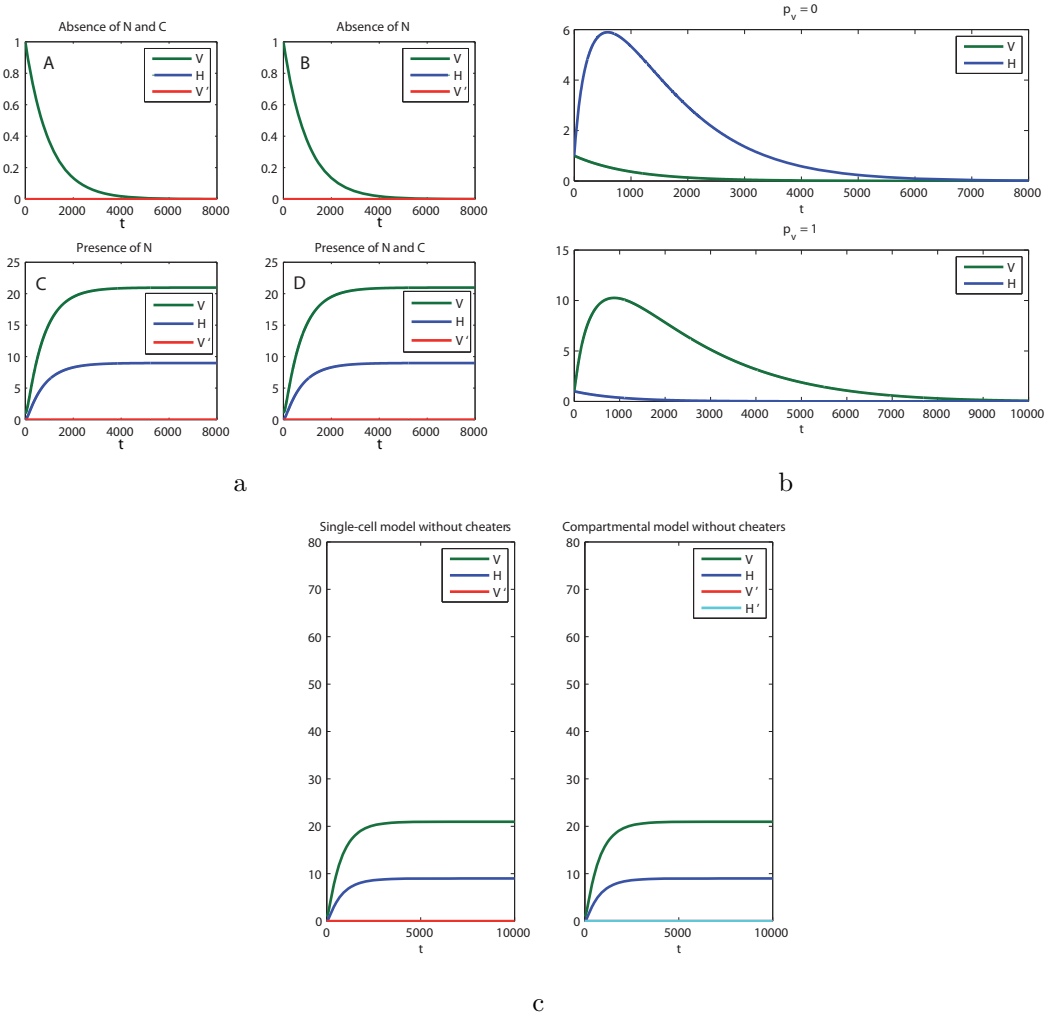
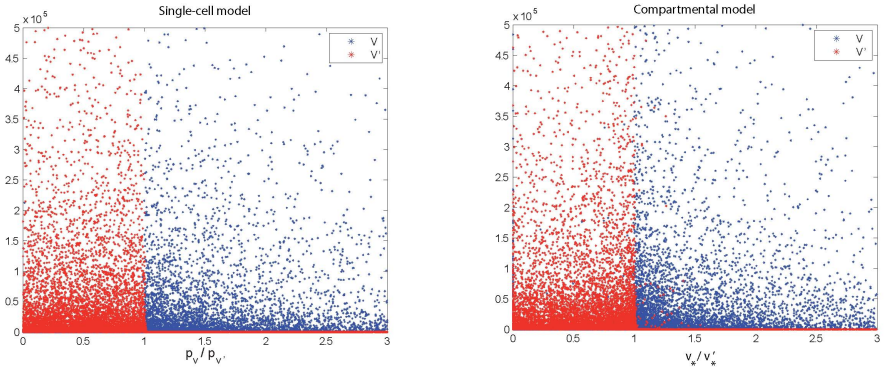


Figure B.1: (a)Role of Initial Resources in the single-cell model. If nitrate is absent in the beginning, the growth of cells is not possible (panels A and B). When it is already present, the system reaches the steady state Y_2 (panels C and D). The result can be generalized to the compartmental model. (b) Vegetative and Heterocyst Solutions for Extreme Values of p_v. The extreme cases in which vegetative cells or heterocysts are not produced lead to the population extinction. (c) Logistic Growth in Absence of Cheaters. In both models, the resident population exhibit a logistic growth toward a steady state.



a b

Figure B.2: (a)Monte Carlo simulation for the homogeneous mix model, with random sampling of all parameters and initial conditions in $\mathbb{R}^{21}$. (b)Monte Carlo simulations for the compartmental model, with random sampling of all parameters and initial conditions in $\mathbb{R}^{22}$.

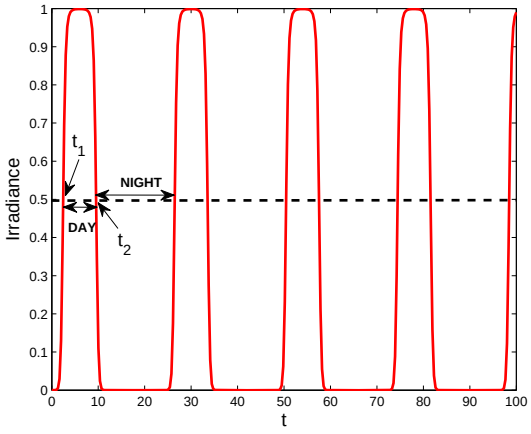


Figure B.3: The function for the irradiance I that represents the alternation of day and night, for $A = 1, \%\text{day} = 0.3, \gamma = 30$. At $t = t_1$ and $t = t_2$, it holds $I(t) = A/2$.

Table B.1: Strains used in this study with accession numbers

Clade and strain	**GenBank accession no.**
Clade I	
Synechococcus sp. PCC 6301	AP008231
Synechococcus sp. CC9605	AY172802
Synechococcus sp. WH8101	AF001480
Prochlorococcus sp. MIT9313	AF053399
Synechocystis sp. PCC6803	NC_000911
Synechocystis sp. PCC6308	AB039001
Synechocystis sp. CR_L29	EF545641
Cyanobium sp. JJ23-1	AM710371
Cyanothece PCC 8801	AF296873
CladeII	
Pleurocapsa sp. CALU 1126	DQ293994
Dermocarpa sp. MBIC10768	AB058287
Myxosarcina sp. PCC 7312	AJ344561
Myxosarcina sp. PCC 7325	AJ344562
Clade III	
Arthrospira PCC 8005	X70769
Lyngbya aestuarii PCC 7419	AB075989
Leptolyngbya antarctica ANT.ACE.1	AY493588
Arthronema gygaxiana	AJ133106
Pseudoanabaena sp. PCC 6802	AB039016
Pseudoanabaena sp. PCC 7403	AB075995
Phormidium mucicola IAM M-221	AB003165
Oscillatoria sancta PCC 7515	AF132933
Oscillatoria sp.	AJ133106
Trichodesmium erythraeum IMS 101	(AF013030)
Spirulina subsalsa IAM M-223	AB003166
Spirulina sp. PCC 6313	X75045
Spirulina laxissima SAG 256.80	DQ393278
Clade IV	
Nostoc sp. PCC 7120	X59559
Nostoc sp. PCC 6720	DQ185240
Anabaena cylindrica PCC 7122	AF247592
Anabaena PCC 7108	AJ133162
Calothrix desertica PCC 7102	AM230699
Calothrix sp. PCC 7103	AM230700
Nodularia sp. PCC 7804	AJ133181
Clade V	
Fischerella muscicola SAG2027	AM709634
Fischerella PCC 7414	AB075986
Chlorogloeopsis fritschii PCC 6912	AB075981
Chlorogloeopsis sp. PCC 7518	X68780
Outgroup	
*Agrobacterium tumefaciens*strain D6	EU155875

C Supplementary material of Chpt. 4

C.1 The irradiance function

The choice of the parameters used in the irradiance function driving the model (equation 5 in the main text) was made to best match the available real data. These data represent the total monthly potential flat surface radiation measured at the center of UTM zones 32N, 32P, 32R, 32T, 32V. In our model, we needed the exact sunrise times at these locations. However, the centers of those UTM zones do not always correspond to a city. It was then easier for us to get the sunrise times of cities at the same latitude of the centers, considering that the longitude does not affect the local sunrise time. Figure SC.1a shows the variation of the monthly irradiance at different latitudes according to the irradiance function of the model. In Figure SC.1b, the comparison between the yearly total irradiance obtained from the model and from the real data indicates that the irradiance function approximate the real irradiance well.

C.2 Analysis of the tradeoff between biomass and population size

According to the results shown in Figures 3-5 in the main text, a high cell investment r corresponds to high levels of biomass production, but not to a large population size (and viceversa). Since the biomass is given by

$$\text{biomass} = \frac{r}{r_d} \cdot V \,,$$

where V is the number of cells, this tradeoff can be generated by a nonlinear change of the number of cells with respect to changes in r. We quantify these variations by computing the sensitivity coefficient $\sigma = \sigma_{V^*}/\sigma_r$. We use the default parameter values listed in Table 1 of the main text and the sample values $a = 0.1, \delta_c = \delta_n = 0.5$ and $I = 1000$ to compute the population size steady state $V^*(r)$ for the circadian species. For these particular values, we use symbolic algebra in Maple to numerically solve for the equilibria of system (1) of the main text. We have that

$$V^*(r) = -0.1 + \frac{60}{r}\,.$$

Given two values r'' and r' where $r'' > r'$, we define

$$\sigma_{V^*} = \frac{\Delta_{V^*}}{V^*(r')} = \frac{V^*(r'') - V^*(r')}{V^*(r')}\,,$$

$$\sigma_r = \frac{\Delta_r}{r'} = \frac{r'' - r'}{r'}\,.$$

$$\sigma = \sigma_{V^*}/\sigma_r\,.$$

The coefficient σ is a measure of the sensitivity of the population steady state to variations in r. Figure C.3 shows the plot of σ vs r, where r ranges between 10^{-5} and 10 as in the main simulations. The graph shows that $-1 < \sigma < 0$ and that σ decreases exponentially with increasing r. As σ approaches -1, the tradeoff becomes weaker. We can then conclude that increases in r lead to a lower sensitivity of population size in respect to r (as evidenced by the decrease in $|\sigma|$).

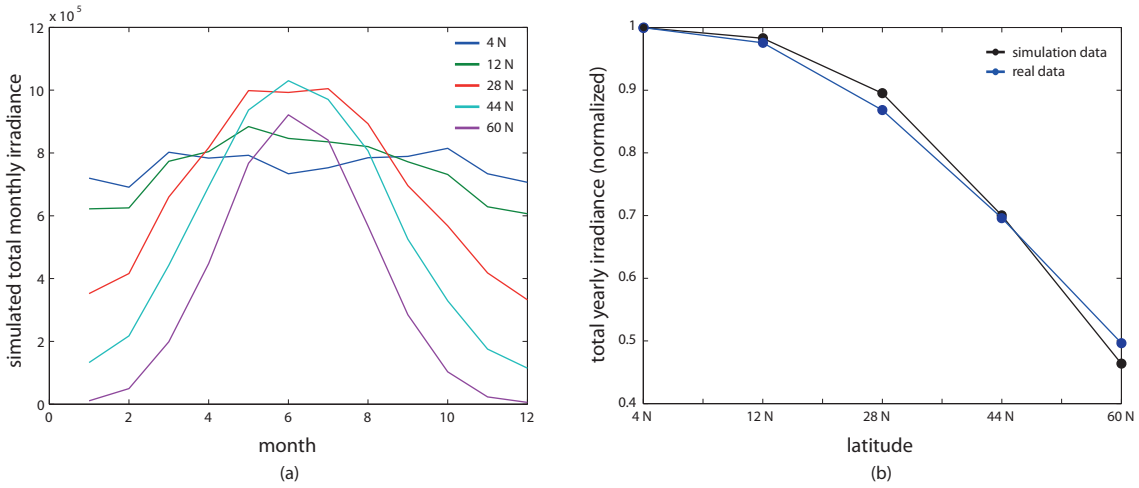


Figure C.1: (a)Total monthly irradiance at the five considered latitudes computed according to the irradiance function used in the model. (b) Normalized difference of the total yearly irradiance at different latitudes based on real data (blue) and model (black).

Table C.1: Value of a corresponding to the highest mean biomass for each r value, latitude and species type.

		r values												
Lat.	Sp.	10^{-5}	$5 \cdot 10^{-5}$	0.0001	0.0005	0.001	0.005	0.01	0.05	0.1	0.5	1	5	10
4N	Circ	0.1	0.5	0.05	0.1	5	0.05	0.05	0.05	0.05	0.1	0.05	5	0.5
	Het	0.1	1	1	5	5	0.05	1	0.1	0.1	1	1	5	5
	Unc	0.1	0.1	0.1	1	1	1	0.05	0.05	0.05	0.1	0.5	1	5
12N	Circ	0.1	0.1	0.5	1	5	0.5	0.05	0.05	0.05	0.1	0.05	1	0.5
	Het	0.1	0.5	1	1	5	0.05	0.05	0.1	0.1	1	1	5	5
	Unc	0.1	0.05	0.1	0.5	1	5	5	0.05	0.1	0.1	0.5	1	5
28N	Circ	0.1	0.05	0.5	0.1	0.05	0.05	0.05	0.05	0.05	0.5	0.05	1	0.5
	Het	0.1	0.5	1	5	5	0.05	0.05	0.1	0.1	1	1	5	5
	Unc	0.05	0.05	0.05	0.5	1	5	0.05	0.05	0.1	0.1	1	1	5
44N	Circ	0.05	1	0.5	0.5	1	0.05	0.05	0.05	0.05	0.1	0.5	1	5
	Het	0.05	0.5	1	5	5	0.05	0.05	0.1	0.1	1	1	5	5
	Unc	0.05	0.05	0.1	1	1	5	0.05	0.05	0.05	0.1	1	1	5
60N	Circ	0.1	0.5	0.1	0.1	5	0.05	0.05	0.1	0.05	0.1	0.5	5	0.05
	Het	0.1	0.5	0.1	1	5	0.05	0.1	0.1	0.1	0.5	1	0.1	5
	Unc	0.05	0.1	0.1	0.5	1	5	0.05	0.1	0.1	0.1	0.1	1	1

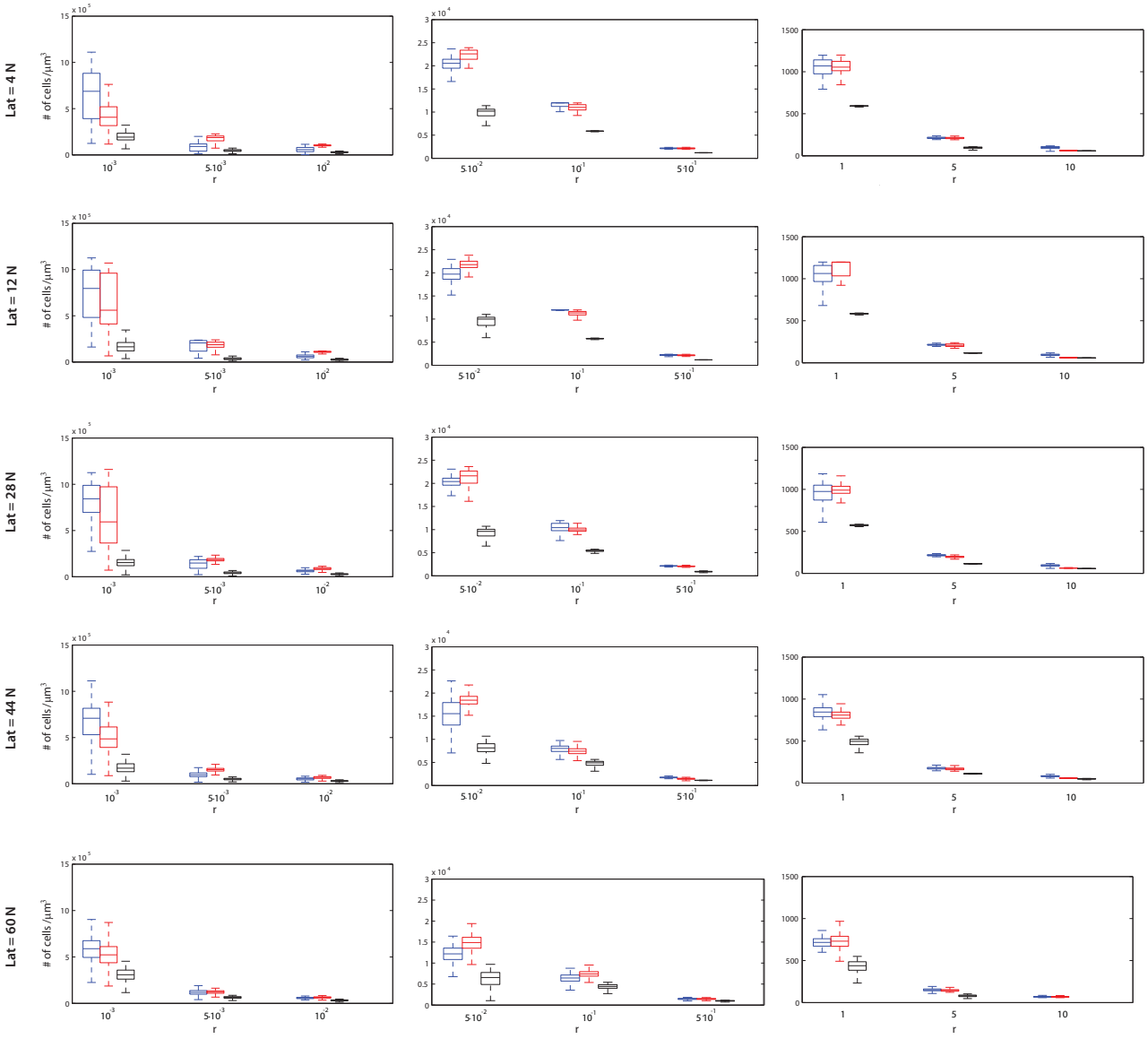


Figure C.2: Zoomed view of Figure 2 of the main text for $r \geq 10^{-3}$. Box plots represent the population size recorded daily over a 35 years period at latitudes 4N, 12N, 28N, 44N and 60N. The value of a used for these box plots provides the maximum mean biomass for the corresponding r and latitude.

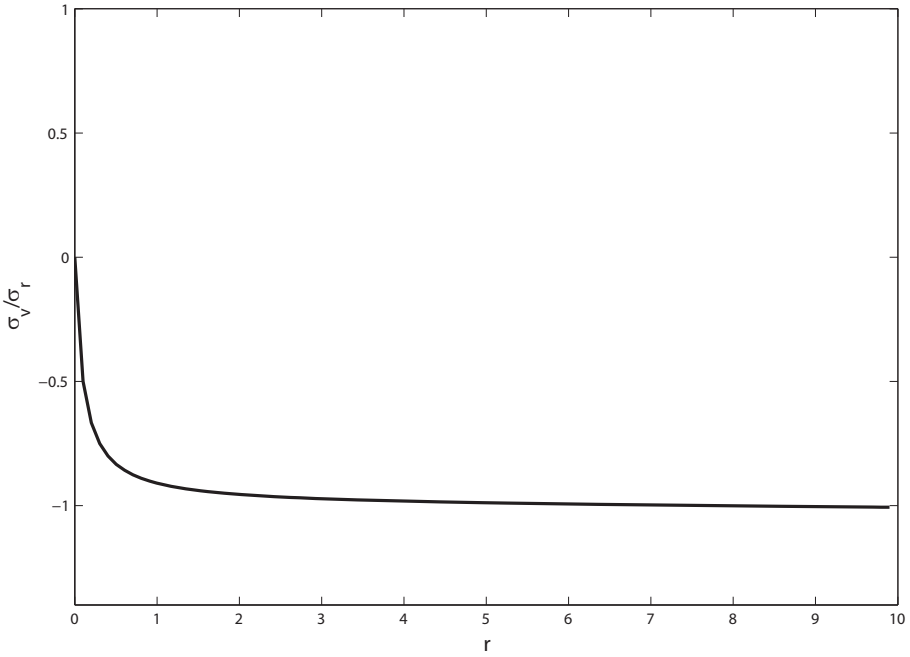


Figure C.3: Sensitivity coefficient σ_{V^*}/σ_r for values of r ranging between 10^{-5} and 10.

D Future work

D.1 Coexistence of carbon and nitrogen fixation in *Trichodesmium*

As already mentioned in previous chapters, cyanobacteria that fix both carbon and nitrogen separate these two processes in time or in space. This is because the oxygen produced during photosynthesis inhibits the functioning of nitrogenase. Temporal division of labor corresponds to circadian species, while spatial division corresponds to heterocystous species. However *Trichodesmium*, a cyanobacterial species, represents a different and unique case. As a future offshoot of this thesis, I aim to study the mechanisms for the coexistence of the fixation processes in *Trichodesmium* by considering the biophysical properties of this species.

Trichodesmium is ubiquitous in oceans at tropical latitudes. Filaments of *Trichodesmium* do not have heterocyst cells, but they fix nitrogen in the daylight period, when photosynthesis is performed and oxygen levels are high. According to several studies, nitrogenase is expressed in clusters of cells (diazocytes) usually situated in the middle of the filament [30, 81, 111, 13, 166]. However, diazocytes do not exhibit special morphological properties as a thick cell wall and are not terminally differentiated. In *Trichodesmium*, nitrogen and carbon fixation do not alternate according to a circadian rhythm, however photosynthesis is lowered around mid-day, when nitrogen fixation reaches its maximum [13, 44, 51]. Although these findings could suggest the presence of reversible differentiation mechanisms, the strategy for the coexistence of carbon and nitrogen fixation is still a matter of debate and not a fully understood problem.

Natural populations of *Trichodesmium* are mainly found in colonies composed of parallel filaments (tufts) or trichomes arranged in star-shaped bundles (puffs).

Single filaments or small colonies are not always able to fix nitrogen [135]. Usually, diazotrophic cells occupy a central position in the filament [13, 44]. However, considering a section of the entire colony, the cells with nitrogenase appear to be more randomly distributed [56]. The spatial arrangement in the colonies might favor the creation of low oxygen regions, where nitrogen fixation can take place successfully. Direct measurements of O_2 concentration in *Trichodesmium* colonies has provided evidence of micro-oxic zones in the bundles [135].

D.2 Open questions

The expression of nitrogenase only in defined, compact regions could favor the balance of gas and nutrients fluxes along the filament, hence representing a way to allow for the coexistence of nitrogen fixation and photosynthesis. If no thick cell wall separates the diazocytes from the other cells, the success of *Trichodesmium* strategy would be related to how nitrogen, oxygen and other chemicals diffuse in the filament and in a colony. The study of dissolved gases diffusion in *Trichodesmium* can reveal whether the clustering of diazocytes plays a role in preventing inhibition due to oxygen. It has been hypothesized that the spatial pattern of cells in colonies creates a micro-oxic region in which nitrogenase is protected from the surrounding oxygen. By numerically simulating gas dynamics in the whole colony, I wish to investigate the extent to which this assumption is valid. In particular, I would like to investigate the role of spatial arrangement of diazocytes and of the ratio between diazocytes and photosynthetic cells on the formation of a low oxygen region in the colony.

D.3 Existing theoretical work on *Trichodesmium*

Although Trichodesmium is widely studied, the topic of gas dynamics in filaments and colonies is rarely approached. Staal et al. [188] developed a reactive-transport model of the influx of oxygen to heterocysts and diazocytes, providing a plausible explanation for the absence of *Trichodesmium* from freshwater and for its abundance in tropical oceans. However, they only considered diffusion into one cell

without taking into account the rest of the filament.

The internal exchange mechanisms and the physiology of filamentous cyanobacteria are still unresolved issues [52, 126]. Walsby hypothesized that, in differentiated filaments, gas exchange takes place mainly through actively regulated terminal pores of the heterocystous cells [210]. Though providing useful descriptions of heterocysts morphology, Walsby's work presently lacks empirical proof. Regarding diffusion in a colony, previous attempts to derive oxygen concentration in a colony by means of theoretical calculations failed in matching the experimental data [31].

D.4 The mathematical model

Reaction-diffusion systems are a suitable mathematical framework for the simulation of gas diffusion in a filament or in a colony. We model a tuft in which we consider three main reactions, namely photosynthesis, nitrogen fixation and cellular respiration. As shown in Figure D.1, the domain representing a tuft consists in a rectangular domain $D = D_1 \cup D_2 \cup D_3$, with boundary ∂D, where $D_i, i = 1, 2, 3$ indicate a domain region in which a particular reaction occurs. For each substance involved in the mentioned reactions, we provide a partial differential equation. The variables of the system are then: oxygen $O(\mathbf{x}, t)$, sugar (fixed carbon) $S(\mathbf{x}, t)$, fixed nitrogen $N(\mathbf{x}, t)$, carbon dioxide $C(\mathbf{x}, t)$ and free atmospheric nitrogen $N_2(\mathbf{x}, t)$. For the ease of reading, in the following we will omit writing the dependence of these variables on $\mathbf{x}$ and t. Oxygen and sugar are produced in the photosynthetic zone (green, D_1), while nitrogen is produced in the diazotrophic area (blue, D_2). Since oxygen and nitrogen seep out of the cells, we also consider a certain amount of empty spaces (gray, D_3) where diffusion occurs without production . Free nitrogen, oxygen and carbon dioxide are assumed to be always present in some concentration in the surrounding water. Considering a time interval $[0, T]$, the set of partial differential equations of the model are:

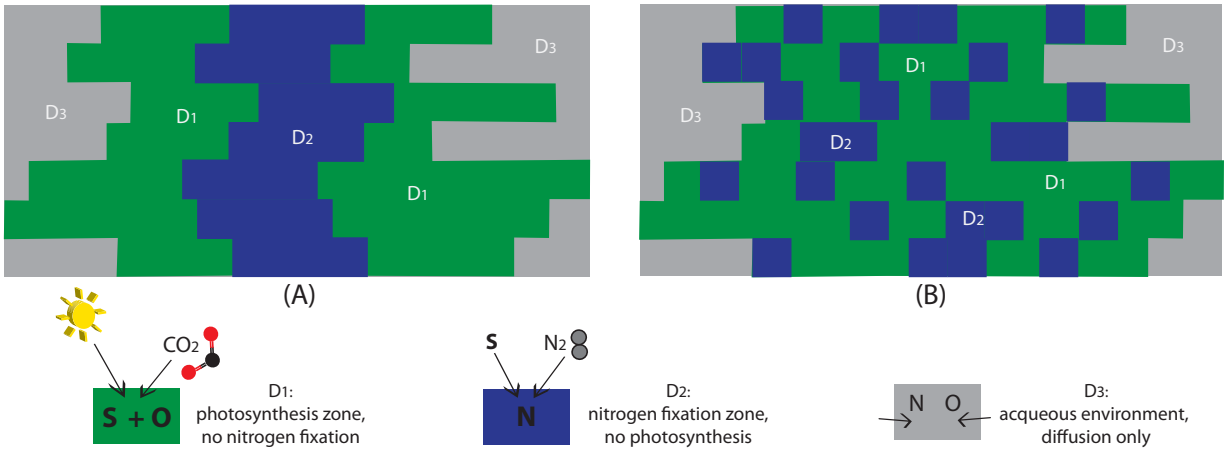


Figure D.1: Examples of schematic domain of a tuft-shaped *Trichodesmium* colony. In (A), diazocytes occupy the central region of filaments. in (B), they are randomly distributed. In both examples, the photosynthetic subdomains (D_1) absorb solar light and dissolved CO_2 to produce sugar (fixed carbon) and oxygen. The diazotrophic region (D_2) takes in dissolved nitrogen and sugar to fix nitrogen. In the empty surrounding space (D_3) nitrogen and oxygen can leak out and diffuse passively.

$$
\begin{aligned}
\frac{\partial O}{\partial t} &= D_O \triangle O - \gamma(\mathbf{x}) h_1 R(O,S) - a_1 \psi(\mathbf{x}) P_N(N_2,S,O) + \varphi(\mathbf{x})\, b_1\, P_O(\overline{I},C) \\
\frac{\partial S}{\partial t} &= D_S \triangle S - \gamma(\mathbf{x}) h_2 R(O,S) - a_2 \psi(\mathbf{x}) P_N(N_2,S,O) + \varphi(\mathbf{x})\, b_2\, P_O(\overline{I},C) \\
\frac{\partial N}{\partial t} &= D_N \triangle N - \gamma(\mathbf{x}) h_N \frac{N}{N+k_N} + \psi(\mathbf{x})\, a_3\, P_N(N_2,S,O) \\
\frac{\partial C}{\partial t} &= D_C \triangle C - \varphi(\mathbf{x})\, b_3\, P_O(\overline{I},C) + \gamma(\mathbf{x}) h_3 R(O,S) \\
\frac{\partial N_2}{\partial t} &= D_{N_2} \triangle N_2 - \psi(\mathbf{x})\, a_4\, P_N(N_2,S,O)
\end{aligned}
\tag{D.1}
$$

where

$$
\begin{aligned}
P_O(I,C) &= \frac{1}{\frac{1}{k_0}+\frac{1}{k_{\bar{I}}\bar{I}}+\frac{1}{k_C C}+\frac{1}{k_{\bar{I}C}\bar{I}C}}\,,\\
P_N(N_2,S,O) &= \frac{1}{1+k_{inib}O}\cdot\frac{1}{\frac{1}{k_0}+\frac{1}{k_{N_2}N_2}+\frac{1}{k_S S}+\frac{1}{k_O O}+\frac{1}{k_{N_2O}N_2O}+\frac{1}{k_{SN_2}SN_2}+\frac{1}{k_{SO}SO}}\\
R(O,S) &= \frac{1}{\frac{1}{k_0}+\frac{1}{k_O O}+\frac{1}{k_S S}+\frac{1}{k_{OS}OS}}\,,
\end{aligned}
\tag{D.2}
$$

and

$$
\begin{aligned}
\gamma(\mathbf{x}) &= \begin{cases} 0 & \text{if} \quad \mathbf{x}\in D_3 \\ 1 & \text{if} \quad \mathbf{x}\in D_1, D_2 \end{cases}\\
\psi(\mathbf{x}) &= \begin{cases} 0 & \text{if} \quad \mathbf{x}\in D_1, D_3 \\ 1 & \text{if} \quad \mathbf{x}\in D_2 \end{cases}\\
\varphi(\mathbf{x}) &= \begin{cases} 0 & \text{if} \quad \mathbf{x}\in D_2, D_3 \\ 1 & \text{if} \quad \mathbf{x}\in D_1 \end{cases}
\end{aligned}
\tag{D.3}
$$

In order for equations (D.1) to be a well-posed problem, we provide the following boundary and initial conditions:

$$
\begin{aligned}
O(\mathbf{x}) = O_{diss} = 7\ ag\ \mu m^{-3} \quad \text{in} \quad \partial D\times[0,T], \qquad O(\mathbf{x},0) = O_0 = 1\ ag\ \mu m^{-3}\\
S(\mathbf{x}) = 0 \quad \text{in} \quad \partial D\times[0,T], \qquad S(\mathbf{x},0) = S_0 = 0\,,\\
N(\mathbf{x}) = 0 \quad \text{in} \quad \partial D\times[0,T], \qquad N(\mathbf{x},0) = N_0 = 0\,,\\
C(\mathbf{x}) = C_{diss} = 90\ ag\ \mu m^{-3} \quad \text{in} \quad \partial D\times[0,T], \qquad C(\mathbf{x},0) = C_0 = 1\ ag\ \mu m^{-3}\\
N_2(\mathbf{x}) = N_{2,diss} = 12.5\ ag\ \mu m^{-3} \quad \text{in} \quad \partial D\times[0,T], \qquad N_2(\mathbf{x},0) = N_{2,0} = 1\ ag\ \mu m^{-3}\,.
\end{aligned}
$$

In system (D.1), the first term of the equations represents diffusion, characterized by the coefficients $D_{(\cdot)}$ specific for each element. The function $R(O,S)$ represents cellular respiration, whereby oxygen and sugar are consumed to produce carbon. $P_N(N_2,S,O)$ corresponds to the production of fixed nitrogen, that needs free nitrogen as row material, and sugar and oxygen for energy. The term $1/(1+k_{inib}O)$

represents the inhibition of nitrogen fixation due to the presence of oxygen. Photosynthesis is represented by the function $P_O(\overline{I}, C)$, whereby the constant solar energy $\overline{I}$ and carbon dioxide are used to produce sugar and oxygen. Water is assumed non-limiting and not included in the equations. Functions R, P_N and P_O are Michaelis-Menten functions for multiple substrates. This class of functions allows the use of all the considered substrates as limiting factors for the output production. The coefficients $k_{(.)}$ are the specificity constant for each substrate and their products. The coefficients $a_{(.)}, b_{(.)}$ and $h_{(.)}$ indicate the rate of consumption or production of the different elements and are can be derived from stoichiometric calculations. Functions $\gamma(\mathbf{x}), \psi(\mathbf{x})$ and $\varphi(\mathbf{x})$ depend on space and indicate the regions of the domain in which the corresponding reaction takes place.

D.5 Simulation plan and scope

The model will be simulated first in 2-D, and according to the results obtained, possibly extended to 3-D. The first step will be to explore distinct tuft domains with different aggregation patterns and proportions of nitrogen fixing and vegetative cells. For each of these domains we will study the performance of the tuft in terms of nitrogen and carbon production. We have already developed an algorithm that given a fixed number of cells in the tuft, generates random domains. We additionally need an algorithmic procedure that evaluates the degree and type of aggregation of the nitrogen fixing cells for any randomly generated tuft pattern.

This first set of simulations is expected to determine the optimal cell distribution in *Trichodesmium* tufts. We will then develop an evolutionary algorithm to study the evolution of a random population of *Trichodesmium* filaments towards the optimal tuft pattern.

By mean of this model we aim to simulate the evolutionary conditions leading *Trichodesmium* to develop the colony structures observed in nature. This would contribute in disentangling the biophysical mechanisms making *Trichodesmium* a very successful species of our oceans.

Acknowledgments

Homayoun, thank you! You gave me the opportunity of pursuing a career in a fascinating research environment and you guided me in exploring the frontiers of scientific knowledge. With your patience and your great expertise, you mentored me in every step of my PhD, sharing every joy for the results achieved and being of fundamental help in the more difficult phases. Your new perspectives and ideas constantly made the research activity absolutely challenging. I sincerely thank you for your supervision, certainly special, as you always took care not only of the scientific aspects, but also of my characteristics as a person, with tactfulness and understanding.

Andrew Barbour, thank you! Your wise advise, your interest in new ideas, your kind availability significantly supported the developments of my work.

Luca Pavarino, thank you! You supported me in undertaking a PhD, and you actively followed the steps of my academic career.

Petros, thank you! You took me on board of your ship and gave me the first tools to orient myself in the big see of research.

Martin Ackermann, Andrew Barbour, Sebastian Bonhoeffer, Andreas Wagner, I sincerely thank you for being part of my PhD committee and for the time and interest that you dedicated to my work.

Manuela, Bettina, Miro, thank you! Collaborating with you and learning from your expertise was a constant source of enrichment. Thank you for your honesty and commitment that made our collaborations enjoyable and fruitful. I specially thank you ***Manuela***, for your friendship, the patient support, the time you dedicated to me, and for all the pleasant italian conversations!
Anni, thank you! Your professionalism and your friendly care for the group simplified and supported significantly my PhD everyday life!

Mamma, Papà, thank you! From the bottom of my heart, I thank you for all the love that you have been giving me: a love that filled any distance, a love that guided my steps, and that was essential to face all the events during these years. Thank you for having been always ready and prompt in helping, supporting and advising me. You are always the safe and unique harbor from which I can take off and come back in the adventure of building my life.
Nonni, thank you! I am truly grateful to you for your support, protection and your always valuable advice.
Zia Pim Pim, thank you! Your brilliant interest and involvement have been extremely important.
Zia Paola e Zio Ambrogio, thank you! You always followed my steps with a special care and love that are for me absolutely precious.

Chiara, Lara, Andrea, thank you! Your friendship kept overcoming geographical distances, being really a fundamental support and encouragement during every phase of my PhD. Thank you for being always available to share the joys and the efforts of my experience.

Daniel Schmid, thank you! Example of rectitude and reliability, you are for me a fundamental point of reference. Through your *vivace* direction of unforgettable musical experiences, you help me to grow up as a musician and in everyday life.
Friends of the Aaargauer Kantorei and Kantorat Grossmünster Zürich, especially ***Marlise*** and ***Susanne***, thank you all! You are for me a big family and a source of joy that colors my days in *Allegro* and *più animato* tones.

Akos, Sorcha, Tugce, Aditya and Shalmali, Daniel and Naomi, Ilona, Filippo, Tobias and Carin, Vangelis and all my friends, thank you! Each one of you donated me something special. The joyful times spent together, the share of our lives abroad, the mutual help and support animated the various phases and situations of these years.

Curriculum vitae

Valentina Rossetti

Personal Data

Surname	ROSSETTI
Name	Valentina
Date of birth	14.06.1982
Place of Birth	Tradate, VA, Italy
Nationality	Italian
Private address	Schaffhauserstrasse 223, CH-8057 Zürich
Work address	Evolutionary Genetics and Theoretical Biology Group, Institute for Evolutionary Biology and Environmental Studies, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich.
Telephone	+41(0)76 216 74 01
Email	vrossett82@gmail.com

Education

2007–2011 **Phd student in Evolutionary Biology**, *Institute of Evolutionary Biology and Environmental Sciences, University of Zurich*, Zurich, Switzerland.
Supervisor: Prof. Homayoun Bagheri

2007 **Research assistant**, *ETH Zurich, Computer Science and Engineering Lab*, Zurich, Switzerland.
Supervisor: Prof. Petros Koumoutsakos

2004–2006 **Master in Applied Mathematics**, *University of Milano*, Milano, Italy.
Supervisor: Prof. Luca Pavarino. Subject: The propagation of electrical signals in neurons: analysis and development of the Hodgkin-Huxley model.

2001–2004 **Bachelor in Applied Mathematics**, *University of Milano*, Milano, Italy.
Supervisor: Prof. Luca Pavarino. Subject: Eigenvalue problems for elliptic partial differential equations.

1997–2001 **Scientific High School Degree**, *Liceo Scientifico "Leonardo da Vinci"*, Gallarate (VA), Italy.

Continuing education

FS 2011	Lösungsorientiertes Stressmanagement, University of Zurich
	Moderations- und Verhandlungstecknichen, University of Zurich
	HTML Grundlagen als nützliches Wissen zur Erstellung von Websiten, University of Zurich
HS 2010	Stimmtraining - Wirkungsvoll präsentieren, University of Zurich
	Wissenschaft öffentlich kommunizieren, University of Zurich

Languages

English	**Full Proficiency**
German	**Advanced**
French	**Advanced**
Italian	**Mother tongue**

Computer skills

Mathematical software	Matlab, Scilab, Maple, R.
Programming Languages	C,C++.
Operating systems	Microsoft Windows, Mac, Linux.

Teaching Experience

HS 2010 + HS 2011	Teaching assistant for the course "Foundations of mathematical modeling and scientific computing in biology", University of Zurich.

Administrative Experience

08.2011 - today	Secretary of the of the Evolutionary Genetics and Theoretical Biology Group, Institute for Evolutionary Biology and Environmental Studies, University of Zürich.

Experience with the public

10.2010	Collaborator at the Kinderuniversität HS 2010
09.2009	Organization and participation at the exhibition "Der Baum des Lebens" for the Darwin Year, Main Station of Zürich.
04.2008	Organization, oral presentation and active participation to the 175-Year Anniversary of the University of Zürich.

Schaffhauserstrasse 223 – CH-8057 Zurich
✉ *vrossett82@gmail.com* • *http://www.ieu.uzh.ch/staff/phd/rossetti.html*

Additional information

Choral singing
1999-2006: Member of the Coro Laudate, Busto Arsizio, VA, Italy.
2007-today: Collaboration with the Kantorat Grossmünster, Zürich.
2010-today: Member of the Aargauer Kantorei, Aarau.

Writing
4th place at the literary award "Il ponte magico", Edition 2003, Comune di Lavena Ponte Tresa, VA, Italy.

Hiking
2010-today: Member of Zürcher and Aargauer Wanderwege.

Grants

10.2010 Elongation of the Forschungskredit of the University of Zurich for the project "The coexistence of nitrogen fixation and oxygenic photosynthesis in Cyanobacteria"

10.2008 Forschungskredit of the University of Zurich for the project "Multicellularity and the evolution of terminal differentiation in cyanobacteria"

Presentations and conferences

06.2011 **Poster**, *"FEMS European Microbiology Conference*, Geneva, Switzerland.

02.2011 **Invited Talk**, *Institute of Mathematics*, University of Milano, Milano, Italy.

02.2010 **Oral presentation**, *"Biology10"*, University of Neuchatel, Neuchatel, Switzerland.

08.2009 **Poster (in absentia)**, *"13th Int. Symposium on Phototrophic Prokaryota"*, Montreal, Canada.

06.2009 **Workshop**, *"Social evolution: from theory to data (and back again)"*, University of Neuchatel, Neuchatel, Switzerland.

02.2009 **Poster**, *"Biology09*, University of Bern, Bern, Switzerland.

07.2008 **Attendance**, *"XX International Congress of Genetics"*, Berlin, Germany.

Publications

[1] V. Rossetti, B. E. Schirrmeister, M. V. Bernasconi, and H. C. Bagheri. The evolutionary path to terminal differentiation and division of labor in cyanobacteria. *Journal of Theoretical Biology*, 262(1):23 – 34, 2010.

[2] V. Rossetti, M. Filippini, M. Svercel, and H. C. Bagheri. Emergent multicellular life cycles in filamentous bacteria owing to density-dependent population dynamics. *Journal of the Royal Society Interface*, 8:1772, 2011.

[3] V. Rossetti and H. C. Bagheri. Division of labor helps overcome biochemical constraints on biomass production in cyanobacteria. Submitted.

[4] J. F. M. Rodrigues, D. J. Rankin, V. Rossetti, A. Wagner, and H. C. Bagheri. Differences in cell division rates drive the evolution of terminal differentiation in microbes. Submitted.

Schaffhauserstrasse 223 – CH-8057 Zurich
✉ *vrossett82@gmail.com* • *http://www.ieu.uzh.ch/staff/phd/rossetti.html*